

## ПРОЕКТИРАНЕ НА ЛОГИЧЕСКА СХЕМА НА СКЛАД С ДАННИ НА ИНТЕЛИГЕНТНА БИЗНЕС СИСТЕМА

Анна Розева  
Лесотехнически университет - София

Направено е описание на интелигентна бизнес система и функционалността, която предоставя. Представена е архитектурата на съвременна платформа за реализация на интелигентна бизнес система. По-специално разглеждането е насочено към основата на платформата – складовете с данни и метаданни. Представена е същността на логическата схема на склада с данни, описани са елементите ѝ и връзките им със структури на физическата база данни. Представена е методика за проектиране на логическа схема на склада с данните като логически обектно-ориентиран модел на бизнеса, съставен от няколко нива на абстракция. Чрез подобна техника се осигурява динамика на справките, независимост от физическите данни и естествено адаптиране към промени в логическия модел на бизнеса. Показана е структурата на склада с метаданните за съхраняване на проектираната логическа схема.

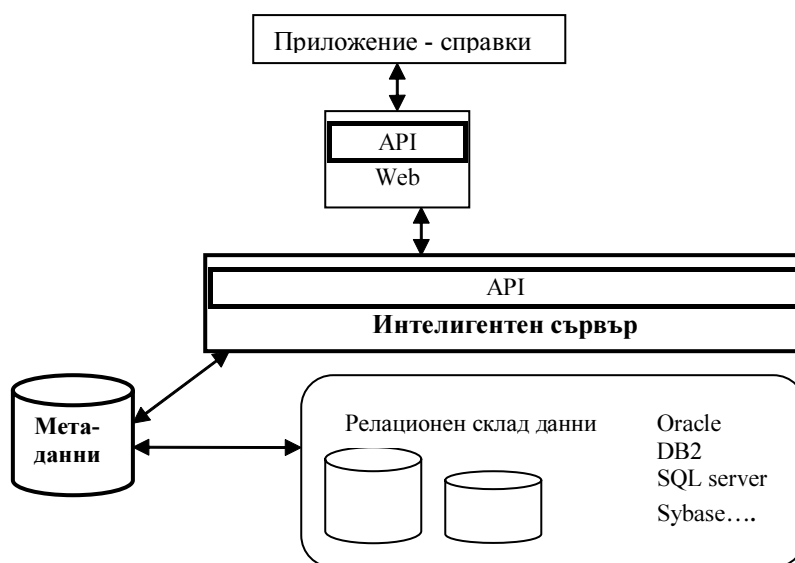
**Ключови думи:** Интелигентна бизнес система, Склад с данни, Логически модел, Логическа схема, Физически модел, Метаданни

**Key words:** Business Intelligence, Data Warehouse, Logical Model, Logical Scheme, Physical Model, Metadata Repository

### Увод

Съвременният електронен и Интернет свят, в който компаниите извършват бизнес, поставя пред информационното им осигуряване предизвикателството да предоставят на потребителите от различни нива интелигентност, вместо както традиционно досега се схващаше просто информация. Разликата се състои в степента на динамичност и контекстното богатство. Вместо показване на статични данни в предварително дефинирани справки, потребностите са насочени все повече към динамично проследяване на тенденции, прогнозиране,

ранжирания, анализ на данните в дълбочина, разработване на “какво - ако” сценарии. Тези и други аналитични приложения осигуряват възможност за пълно вникване в същността на функциониране на бизнеса. Платформата на съвременна интелигентна бизнес система включва един или повече сървъри базирани в Интернет, предоставящи на потребителите разнообразни аналитични услуги и бърз, интуитивен и филтриран достъп до огромни масиви данни, съхранени в различни типове системи [1]. Архитектура на интелигентна бизнес платформа [2] е показана на фиг. 1.



Фиг. 1. Платформа на интелигентна бизнес система  
Fig. 1. Business Intelligence Platform

Особеност на представената платформена архитектура е, че тя е:

- Сървър-базирана – цялата обработка се осъществява от сървъра като клиентът през браузър или (настолна) desktop система има достъп до аналитичната функционалност на платформата;

- Отворена – предлаганата функционалност е достъпна чрез приложни програмируеми интерфейси (API) и по такъв начин интелигентност може да бъде вградена във всякакви приложения;

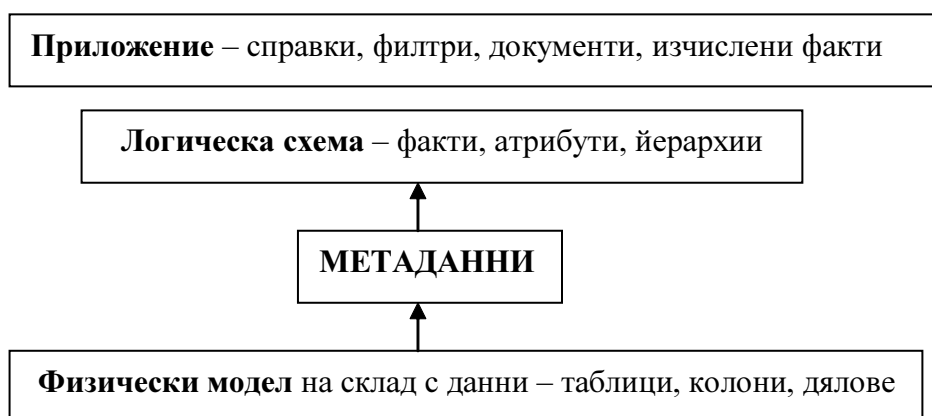
- Прозрачна – поддържа метаданни и изгледи, които скриват сложната физическа организация на данните в базите и позволяват търсенето на информация да става с използване на бизнес термини и правила.

Предмет на разглеждане по-нататък изграждането на логически модел на бизнеса, наречен схема, в качеството ѝ на основа на интелигентната бизнес система – структура на схемата, дефиниране на елементите ѝ, проектиране на връзката им с обектите на физическата база данни и начин на съхраняване чрез метаданни.

#### Логическа схема на склад с данни

Основното предназначение на интелигентна-

та система е да предоставя отговори на обичайни бизнес въпроси под форма на справки. Бизнес въпросите често съдържат логика, която предполага изпълнение на сложни заявки и изчисления над големи обеми от данни. Самите данни обикновено са организирани в релационни складове, които съдържат множество таблици с голям брой колони и трудни за проследяване връзки между тях. Формулирането на заявки за отговор на обичайни управленски въпроси директно върху елементите на склада с данни предполага добро познаване на структурата му и начина на физическа организация и по-често създава практически неудобства. Елиминиране на пряката връзка на структурите данни в базите със заявките се осъществява чрез абстрахиране на физическата база в логически модел на бизнеса. Логическата схема на склада с данни съдържа описанието на логическия модел. Тя служи за опосредстване на връзката между приложенията генериращи справки и физическата база данни като представя структурата и потока на данните в бизнес среда и се съхранява чрез метаданни – фиг.2.



Фиг.2. Метаданните – Абстракция на приложенията от физическия модел данните

Fig.2. Metadata – abstraction of applications from the physical data model

Логическата схема е средството за организация на фактите на бизнес дейността, така че те да могат да бъдат анализирани от различни гледни точки. Компонентите на логическата схема на бизнес данни са [3]:

- **Факти;**

Фактите отразяват уникална ситуация в бизнес дейността, свързана с определена комбинация от атрибути. Те “измерват” бизнеса и са подходящи за обобщаване, напр. продажби, складови наличности, баланси по сметки.

- **Атрибути** (Гледни точки);

Атрибутите поставят фактите в определен

контекст, изискван от справки относно тях. Напр. за факт продажба контекстът се определя от атрибутите продаван продукт, клиент, момент на продажба. Освен контекст атрибутите предоставят и нива за категоризиране и агрегиране на фактите. Така те осигуряват използваемост на фактите в справките на различно ниво на детайлност. Напр. наличието на атрибути ден, месец, тримесечие дава възможност за генериране на справки за дневни, месечни, тримесечни продажби.

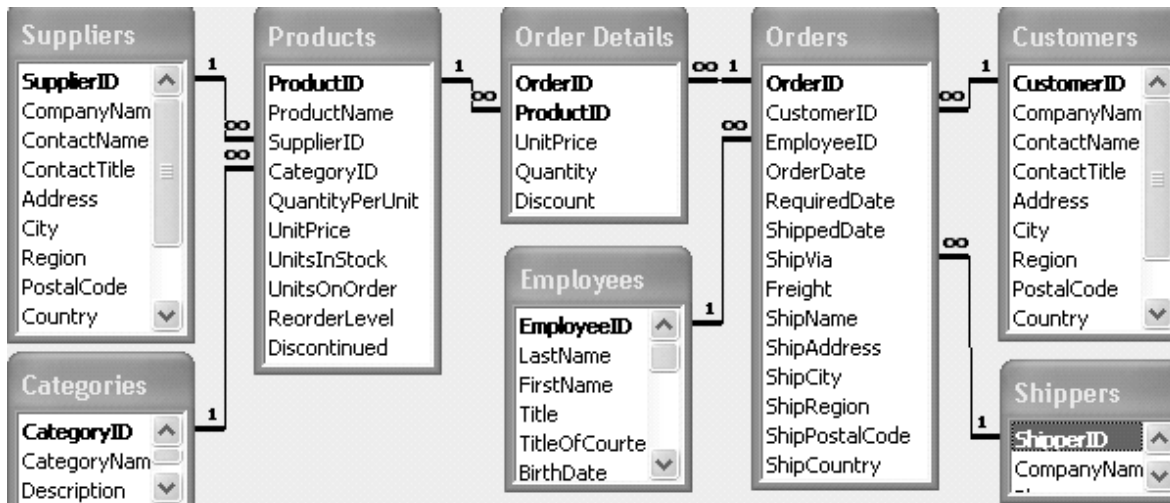
- **Йерархии.**

Йерархиите представляват подредба на атрибути, която отразява връзките между тях от по-аг-

регирано към по-детайлно ниво. Атрибутите от различни йерархии се свързват един с друг посредством факти.

Всеки от обектите на логическата схема представя определени данни от физическата база данни. Физическият модел се изобразява графич-

но чрез диаграма на същностите и връзките между тях (ER диаграма). Примерна (ER диаграма) на релационен модел, представен чрез съвкупност от нормализирани таблици, за описание на бизнес, свързан с продажби [4] е представена на фиг.3.



Фиг.3. Физически модел на база данни от релационен тип  
Fig.3. Physical model of a relational database

Съответствието на елементите на логическата схема с таблиците и колоните в склада с данни е следната:

- Фактът е колона в таблица, която съдържа числови данни. Факт може да се получи също и чрез изчисление над колона (напр. средни/сумарни Продажби) или над няколко колони (напр. Количество\*Цена);
- Атрибут е колона в таблица, съдържаща изключително текстови данни (lookup таблица);
- Йерархията е описание на връзките между атрибути, обикновено 1:M.

#### Проектиране на логическата схема на склад с данни

Проектирането на логическата схема се извършва на основата на ER диаграмата на физическата структура на склада с данни, показана на фиг.3. Това е процес на превръщане на транзакционните данни във вид подходящ за анализи.

##### 1. Дефиниране на факти

Фактите представляват измерители на бизнеса. Уточняването им е първостепенна задача на проектирането. Източник на фактите са колони, които съдържат числови данни (*Data Type Number/Currency*) и не са ключови. Освен това фактите могат да бъдат базови или изчислени.

- Дефиниране на базови факти

За показаната на фиг.4. таблица *Order Details*

числови колони са *UnitPrice* и *Quantity*. За пълноценното уточняване на факта е необходимо да се определи нивото, на което той е регистриран. Факт *Quantity* е регистриран на нива продукт, дата, клиент, служител и превозвач. За последните три нива това се вижда от фиг.3. където таблица *Order Details* е свързана с *Orders* и чрез нея и с таблици *Customers*, *Employees*, *Shippers*. Датата е ниво за факта поради това, че *OrderDate* принадлежи на *Orders*, която е свързана с *Order Details*. Продуктът е ниво, тъй като *ProductID* в *Order Details* е външен ключ на таблица *Product*. На тази база е съставна схемата в дясната част на фиг.4.

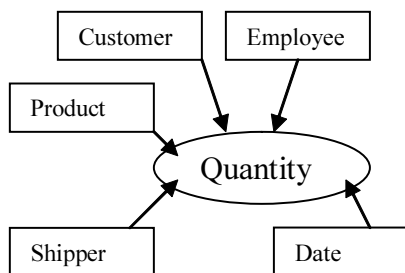
Други факти, произлизащи от колони в таблиците на ER диаграмата са: *UnitPrice*, *Discount* и *Freight* със същите нива като тези на *Quantity*.

- Дефиниране на изчислени факти от базови

Изчислен факт е приход от продажба на продукт. Дефинира се като  $eSales = Quantity * UnitPrice$ . Средната такса превоз за продукт също се получава чрез изчисление:  $eProd\_Freight = (Freight / Sum(Quantity)) * Quantity$ . Крайната цена на продукт се определя като  $ePrice = UnitPrice - (UnitPrice * Discount)$ .

Нивата, на които се регистрират изчислените факти следват от нивата, свързани с базовите и са както на схемата на фиг.5. Таблица 1 съдържа описанията на фактите в логическата схема.

Order Details : Table		
	Field Name	Data Type
🔑	OrderID	Number
🔑	ProductID	Number
▶	UnitPrice	Currency
	Quantity	Number
	Discount	Number



Фиг.4. Дефиниране на факти от таблиците на ER диаграмата  
 Fig.4. Facts' definition from ER diagram tables

Таблица 1

Факт	Нива	Колона
Quantity	Product, Customer, Employee, Shipper, Date	Quantity
UnitPrice	Product, Customer, Employee, Shipper, Date	UnitPrice
Freight	Customer, Employee, Shipper, Date	Freight

**2. Дефиниране на атрибути**

Определянето на основните атрибути на схемата следва от дефинициите на фактите. Атрибутите дефинират гледни точки към тях. Така всяко ниво на факт представлява атрибут на схемата. Следователно атрибути са: *продукт, клиент, служител, превозвач, дата*. Освен тези могат да бъдат дефинирани и други, които да направят възможен анализа на фактите и на други по-обобщени нива. Наричат се категорийни атрибути. Те произлизат от текстови колони в таблици, от които вече са дефинирани основни атрибути. Характерно е, че са с малък брой различни стойности. Такъв атрибут може да е *категорията* продукт. Категорията създава групиране на продуктите и възможност за обобщаване на фактите, регистрирани на ниво про-

дукт. Категорийни атрибути, свързани с атрибутите клиент, служител и превозвач са: *град, регион, държава*, тъй като във всяка от съответните таблици съществуват такива колони. Що се отнася до атрибута дата в анализите на фактите ще е много вероятно да се изисква освен дневно така и месечно, тримесечно и дори годишно ниво. Така *месец, тримесечие и година* следва да се дефинират също като атрибути на схемата.

На атрибутите, които произлизат от колони на една и съща таблица са присъщи връзки. Доколкото атрибут с по-малък брой различни стойности категоризира и групира атрибут с уникални стойности връзката между двата е от тип 1:M. В таблица 2 е показано описанието на част от атрибутите на схемата с връзките между тях.

Таблица 2

Атрибут	Форми	Тип данни	Колона	Връзка			
				Наследник	Родител	Тип	Таблица
Product	ID	Number	ProductID		Category	1:M	Categories
	Name	Text	ProductName		Supplier	1:M	Suppliers
Category	ID	Number	CategoryId	Product		1:M	Products
	Name	Text	CategoryName				
Customer	Id	Number	CustomerId		City	1:M	Customers
	Company	Text	CompanyName				
City	City	Text	City	Customer	Country	1:M	Customers
.....							

### 3. Дефиниране на йерархии

Йерархията в смисъла на логическа схема на склад с данни е логическа подредба на атрибути в гледни точки (фокуси) към предмета на дейност на бизнеса. Гледните точки произтичат от основните атрибути на схемата. Основен атрибут и свързаните с него категорийни оформят йерархия. За настоящата схема могат да се дефинират следните йерархии: Category – Product; Country – City – CustomerName; , Country – City - LastName, Year

– Quarter – Month - Date.

#### Структура на склада с метаданни

Складът с метаданните [5] се състои от релационни таблици. В тях съгласно фиг.2.се съхранява съдържанието на логическата схема и обектите на изгражданите върху нея приложения. Видът и структурата на необходимите таблици, които да съдържат описанията на обектите на логическата схема на склад с данни са следните:

**Таблица за описание на обектите – тип, име, описание, родител, дата на създаване**

OBJECT_ID	OBJECT_TYPE	Type_NAME	OBJECT_NAME	DESCRIPTION	PARENT_ID	CREATE_TIME
0782483A11D4	4	Fact	Quantity		4C0516A211D3	-01-02 20:43:08
078248EE11D4	4	Fact	UnitPrice		4C0516A211D3	-01-02 20:42:52
5553489411D5	8	Warehouse Obj	Tables		5553488011D5	-02-15 20:15:55
5553489511D5	8	Schema Object	Facts		5553488011D5	-02-15 20:16:05
5553489611D5	8	Schema Object	Attributes		5553488011D5	-02-15 20:16:11
6568925D11D5	12	Schema Object	Hierarchy		0E6BA06511D5	-02-15 20:16:16
8D67910711D3	53	Table	CUSTOMERS		8D678D5511D3	-01-02 20:48:55
C9281CA64A7	55	Hierarchy	Time Hierarchy	The Time Hierarchy	2B406BB9469C	-02-05 18:10:50
8D679D3A11D	12	Attribute	Customer City		B6E7CF2711D4	-01-02 20:48:08
8D679D3C11D	12	Attribute	Customer		B6E7CF2711D4	-01-02 20:48:09

**Таблица с дефиниции на обектите – обект, тип, дефиниция**

	Field Name	Data Type
🔑	OBJECT_ID	Text
🔑	DEFINITION_SEQ	Number
	OBJECT_TYPE	Number
	DEFINITION	Memo

**Таблица със свойства на обектите – обект, тип, свойство, стойност, тип на свойството**

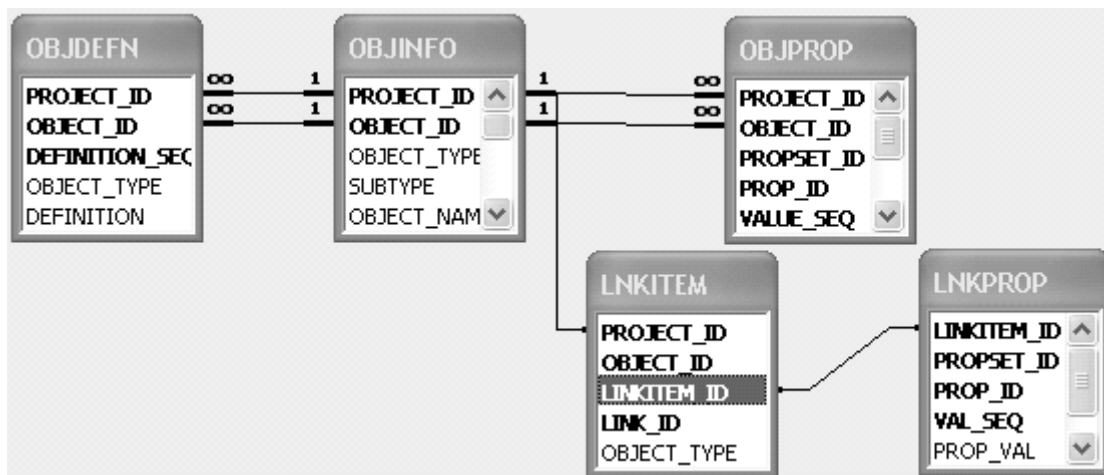
OBJECT_ID	PROPSET_ID	PROP_ID	OBJECT_TYPE	PROP_VAL	PROP_TYPE
11597D4B0E26	5FDF420CBED	4	57	SELECT name FROM sysobjects WHERE name LIKE '#P#%'	8
11597D480E26	5FDF420CBED	6	57	SELECT DatabaseName NAME_SPACE, TableName TAB_NAME, IndexName INDEX_NAME FROM INFORMATION_SCHEMA.TABLES WHERE TABLE_TYPE = 'VIEW'	8
11597D4C0E26	5FDF420CBED	4	57	SELECT name FROM rbw_tables WHERE name LIKE '#P#%' AND create_date <= '2000-01-01'	8
78288B34EE34	A2185AB2DD7	1	57	3	3
78288B34EE34	A2185AB2DD7	13	57	NOT NULL	8
78288B34EE34	A2185AB2DD7	35	57	#primary key #1	8
78288B34EE34	A2185AB5DD7	8	57	mm/dd/yyyy hh:nn:ss	8

**Таблица на връзките – обект, тип, връзка, елемент на връзката**

	Field Name	Data Type
🔑	OBJECT_ID	Text
🔑	LINKITEM_ID	Text
🔑	LINK_ID	Text
	OBJECT_TYPE	Number

**Таблица със свойства на връзките – елемент на връзката, набор свойства, свойство, поредност на стойността, стойност, тип**

	Field Name	Data Type
🔑	LINKITEM_ID	Text
🔑	PROPSET_ID	Text
🔑	PROP_ID	Number
🔑	VAL_SEQ	Number
	PROP_VAL	Memo
	PROP_TYPE	Number



Фиг. 5. Схема на склад с метаданни  
Fig.5. Scheme of a metadata warehouse

Схемата на склада с метаданните е показана на фиг.5.

#### Заклучение

Интелигентните бизнес системи предоставят възможност за генериране на сложни справки над данните от склада с данни на бизнеса. Изключително важно е обектите, с които се конструира съдържанието на справките да са независими от физическата структура на данните. Така справките придобиват динамичен характер и в тях автоматично се отразяват настъпили промени в данните. Средството, което реализира подобна независимост е логическата схема с дефиниции на бизнес правила и мерки и връзката им с колони от таблици с данни. Така логическата схема е схващане за данните от склада по начин, определен от функционирането на бизнеса. Показани са елементите на схемата и начинът на дефинирането им. Описанието на логическата схема като свързващо звено между справките в приложенията и физичес-

ките данни се съхранява в база с метаданни. Базата съдържа релационни таблици с информация за вида, свойствата и връзките между елементите на схемата и таблиците и колоните на склада с данни. Показан е видът на таблиците и съставлящите ги колони, както и схемата на базата с метаданни.

#### Литература

1. Eckerson, W., Next-Generation Business intelligence, What Works. Volume 10. The Data Warehousing Institute.  
<http://www.dw-institute.com/research/display.asp?id=5200>
2. An Architecture for Next-Generation BUSINESS INTELLIGENCE, A Review of the MicroStrategy 7i Architecture for Query & Reporting, Analysis, and Information Delivery, <http://microstrategy.com>
3. Introduction to Microstrategy 7i, Evaluation Guide, Microstrategy Inc, 2003
4. Microsoft T Access 2002, Microsoft Corporation 1992-2001
5. Data Management Association, <http://dama.org>

## LOGICAL SCHEME DESIGN FOR A BUSINESS INTELLIGENCE WAREHOUSE

Anna Rozeva  
University of forestry – Sofia, Bulgaria

### ABSTRACT

A description of business intelligence system and its functionality have been presented. The architecture of a platform for business intelligence system implementation has been outlined. The investigation has been dedicated to the data warehouse and the metadata repository as a foundation of the platform. The basics of the data warehouse's logical scheme have been highlighted – elements and mappings to physical database structures. A methodology for data warehouse scheme design has been presented as a logical object-oriented model of the business having several levels of abstraction. This design technique provides for applications' dynamics, independency from physical data and natural adaptability to changes in the business logical model. A metadata repository structure for storing the logical scheme has been presented.