

- Основными составляющими структуры хранилищ данных являются таблица фактов (fact table) и таблицы измерений (dimension tables).

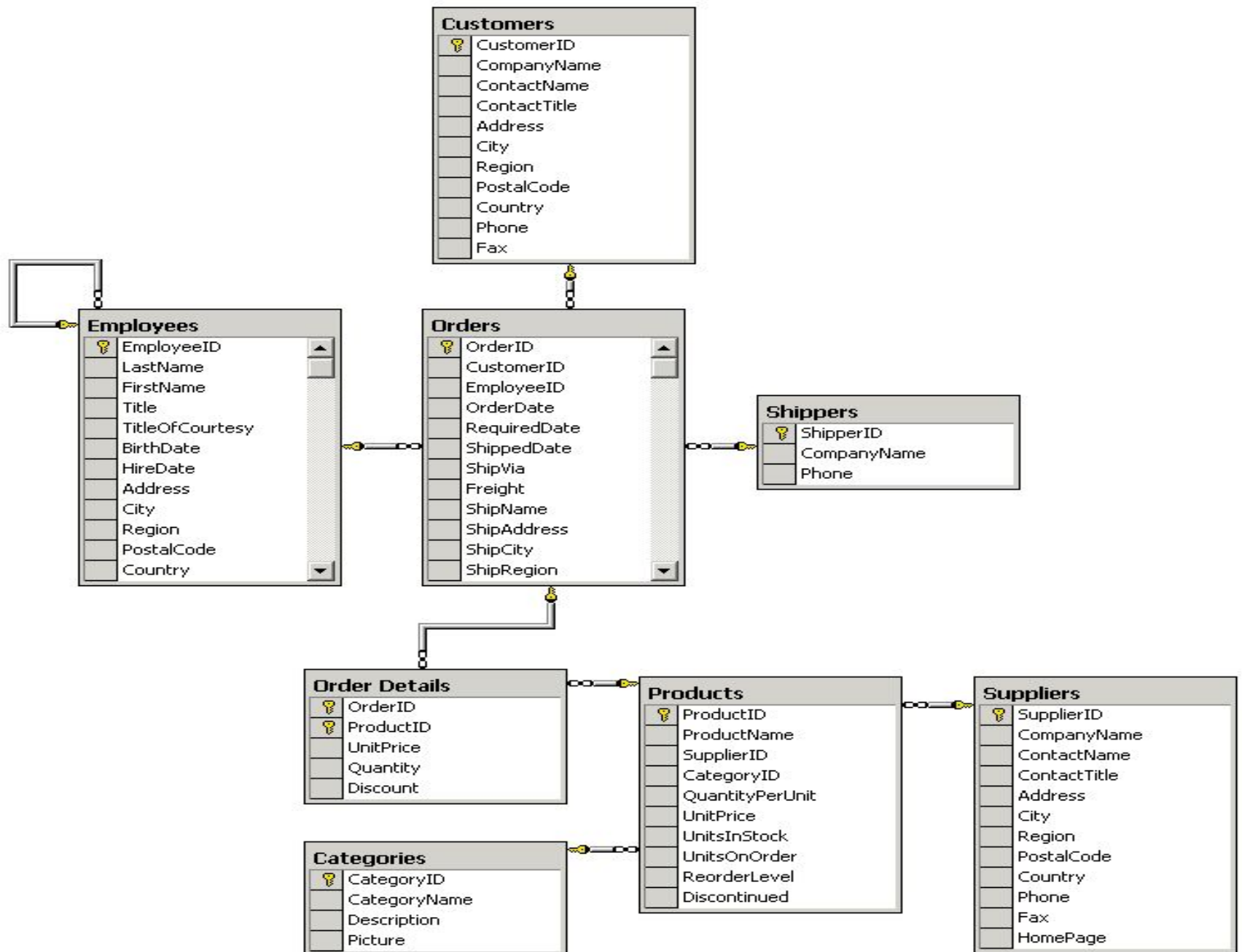


Таблица фактов

- Таблица фактов является основной таблицей хранилища данных. Как правило, она содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться. Обычно говорят о четырех наиболее часто встречающихся **типах фактов**

- факты, связанные с транзакциями (Transaction facts). Они основаны на отдельных событиях (типичными примерами которых являются телефонный звонок или снятие денег со счета с помощью банкомата);
- факты, связанные с «моментальными снимками» (Snapshot facts). Основаны на состоянии объекта (например, банковского счета) в определенные моменты времени, например на конец дня или месяца. Типичными примерами таких фактов являются объем продаж за день или дневная выручка;

- факты, связанные с элементами документа (Line-item facts). Основаны на том или ином документе (например, счете за товар или услуги) и содержат подробную информацию об элементах этого документа (например, количестве, цене, проценте скидки);
- факты, связанные с событиями или состоянием объекта (Event or state facts). Представляют возникновение события без подробностей о нем (например, просто факт продажи или факт отсутствия таковой без иных подробностей).



	TimeKey	CustomerKey	ShipperKey	ProductKey	EmployeeKey	RequiredDate	LineItemFreight	LineItemTotal	LineItemQuantity	LineItemDiscount
5	85	4	11	5	01.08.1996	14.3904	168	12	0	
5	85	4	42	5	01.08.1996	11.992	98	10	0	
5	85	4	72	5	01.08.1996	5.996	174	5	0	
1	79	1	14	6	16.08.1996	2.1321	167.4	9	0	
1	79	1	51	6	16.08.1996	9.476	1696	40	0	
3	34	2	41	4	05.08.1996	10.971	77	10	0	
3	34	2	51	4	05.08.1996	38.3985	1484	35	222.6	
3	34	2	65	4	05.08.1996	16.4565	252	15	37.8	
4	84	1	22	3	05.08.1996	6.0492	100.8	6	5.04	
4	84	1	57	3	05.08.1996	15.123	234	15	11.7	
4	84	1	65	3	05.08.1996	20.164	336	20	0	
2	76	2	20	4	06.08.1996	19.54	2592	40	129.6	
2	76	2	33	4	06.08.1996	12.2125	50	25	2.5	
2	76	2	60	4	06.08.1996	19.54	1088	40	0	
5	34	2	31	3	24.07.1996	11.404	200	20	0	

Таблицы измерений

- Таблицы измерений содержат неизменяемые либо редко изменяемые данные. В подавляющем большинстве случаев эти данные представляют собой по одной записи для каждого члена нижнего уровня иерархии в измерении. Таблицы измерений также содержат как минимум одно описательное поле (обычно с именем члена измерения) и, как правило, целочисленное ключевое поле (обычно это суррогатный ключ¹) для однозначной идентификации члена измерения.

SQL Server Enterprise Manager - [4:Data in Table 'Product_Dim' in 'Northwind_Mart' on 'MAINDESK']

Console Window Help

SQL

	ProductKey	ProductID	ProductName	SupplierName	CategoryName	ListUnitPrice
▶	1	1	Chai	Exotic Liquids	Beverages	18
	2	2	Chang	Exotic Liquids	Beverages	19
	3	3	Aniseed Syrup	Exotic Liquids	Condiments	10
	4	4	Chef Anton's Cajun Seasoning	New Orleans Cajun Delights	Condiments	22
	5	5	Chef Anton's Gumbo Mix	New Orleans Cajun Delights	Condiments	21.35
	6	6	Grandma's Boysenberry Spread	Grandma Kelly's Homestead	Condiments	25
	7	7	Uncle Bob's Organic Dried Pears	Grandma Kelly's Homestead	Produce	30
	8	8	Northwoods Cranberry Sauce	Grandma Kelly's Homestead	Condiments	40
	9	9	Mishi Kobe Niku	Tokyo Traders	Meat/Poultry	97
	10	10	Ikura	Tokyo Traders	Seafood	31
	11	11	Queso Cabrales	Cooperativa de Quesos 'Las Cabras'	Dairy Products	21
	12	12	Queso Manchego La Pastora	Cooperativa de Quesos 'Las Cabras'	Dairy Products	38
	13	13	Konbu	Mayumi's	Seafood	6
	14	14	Tofu	Mayumi's	Produce	23.25
	15	15	Genen Shouyu	Mayumi's	Condiments	15.5
	16	16	Pavlova	Pavlova, Ltd.	Confections	17.45
	17	17	Alice Mutton	Pavlova, Ltd.	Meat/Poultry	39
	18	18	Carnarvon Tigers	Pavlova, Ltd.	Seafood	62.5
	19	19	Teatime Chocolate Biscuits	Specialty Biscuits, Ltd.	Confections	9.2

Схема «звезда»

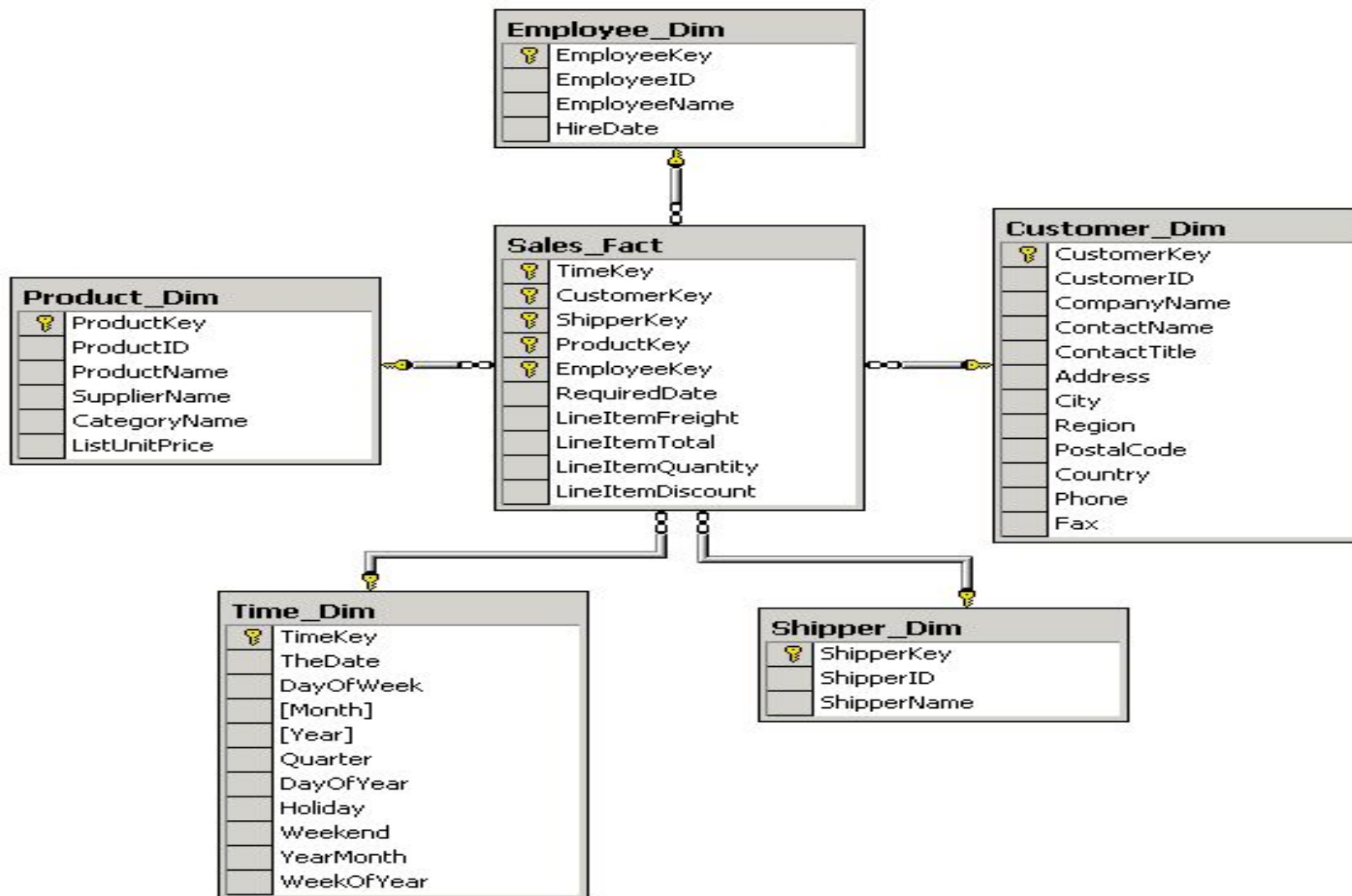


Схема «снежинка»

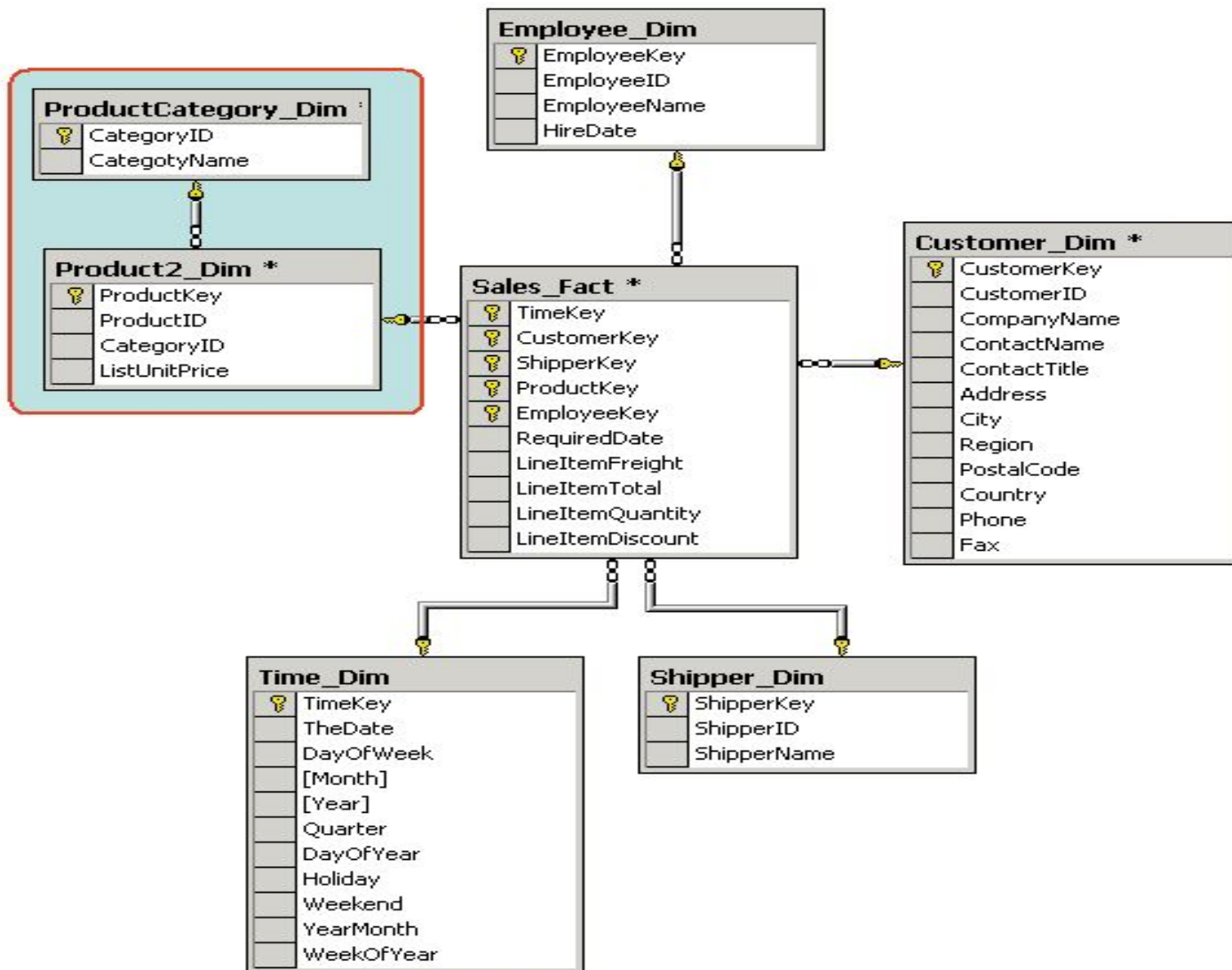
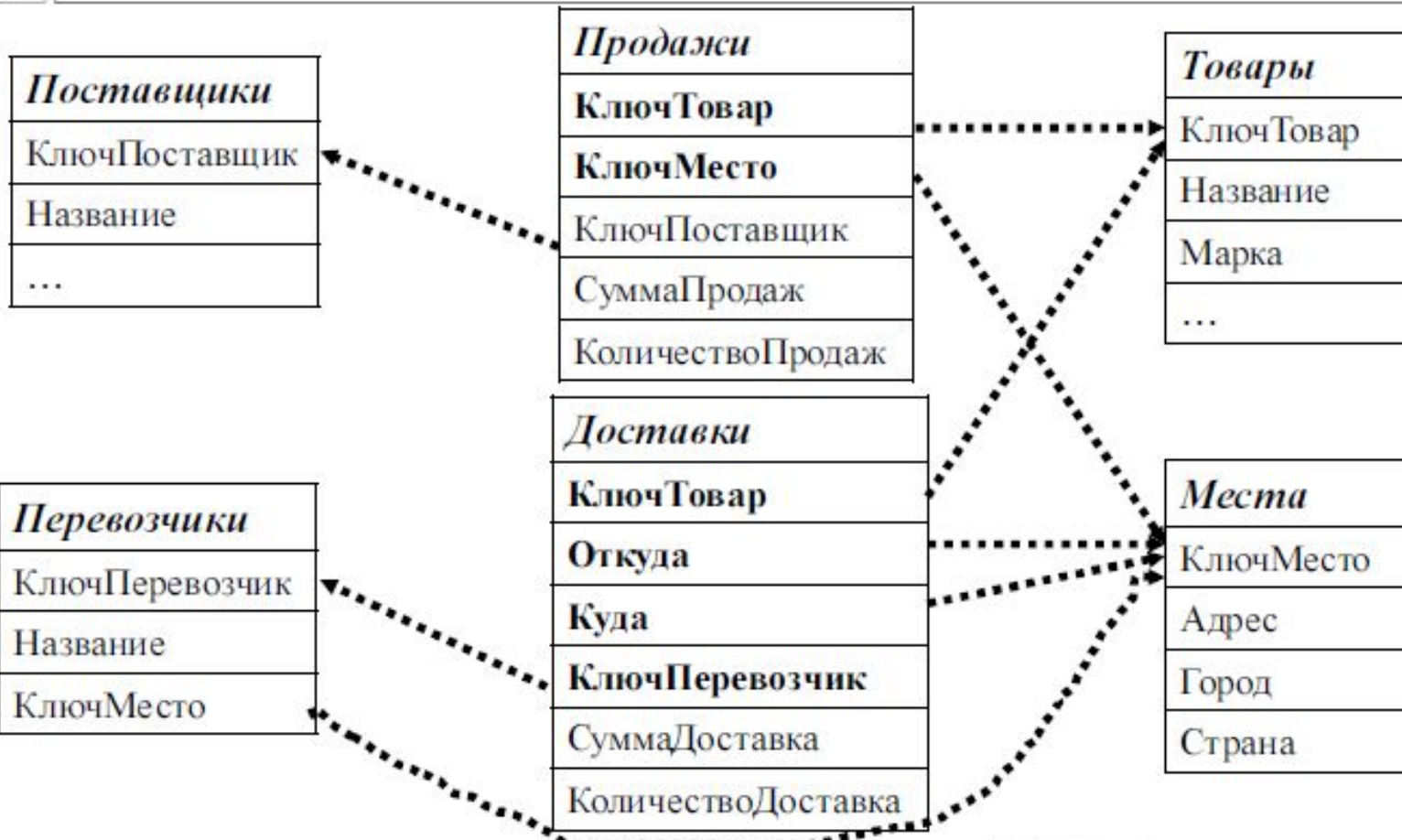


Схема "созвездие"

20



Проектирование хранилища: 4 точки зрения

23

- Взгляд сверху-вниз
 - какая информация является релевантной и необходимой для хранилища?
- Источники данных
 - какая информация из источников данных будет помещаться в хранилище?
- Хранение данных
 - какими таблицами измерений и таблицами фактов представлено хранилище?
- Бизнес
 - какая информация требуется конечному пользователю?

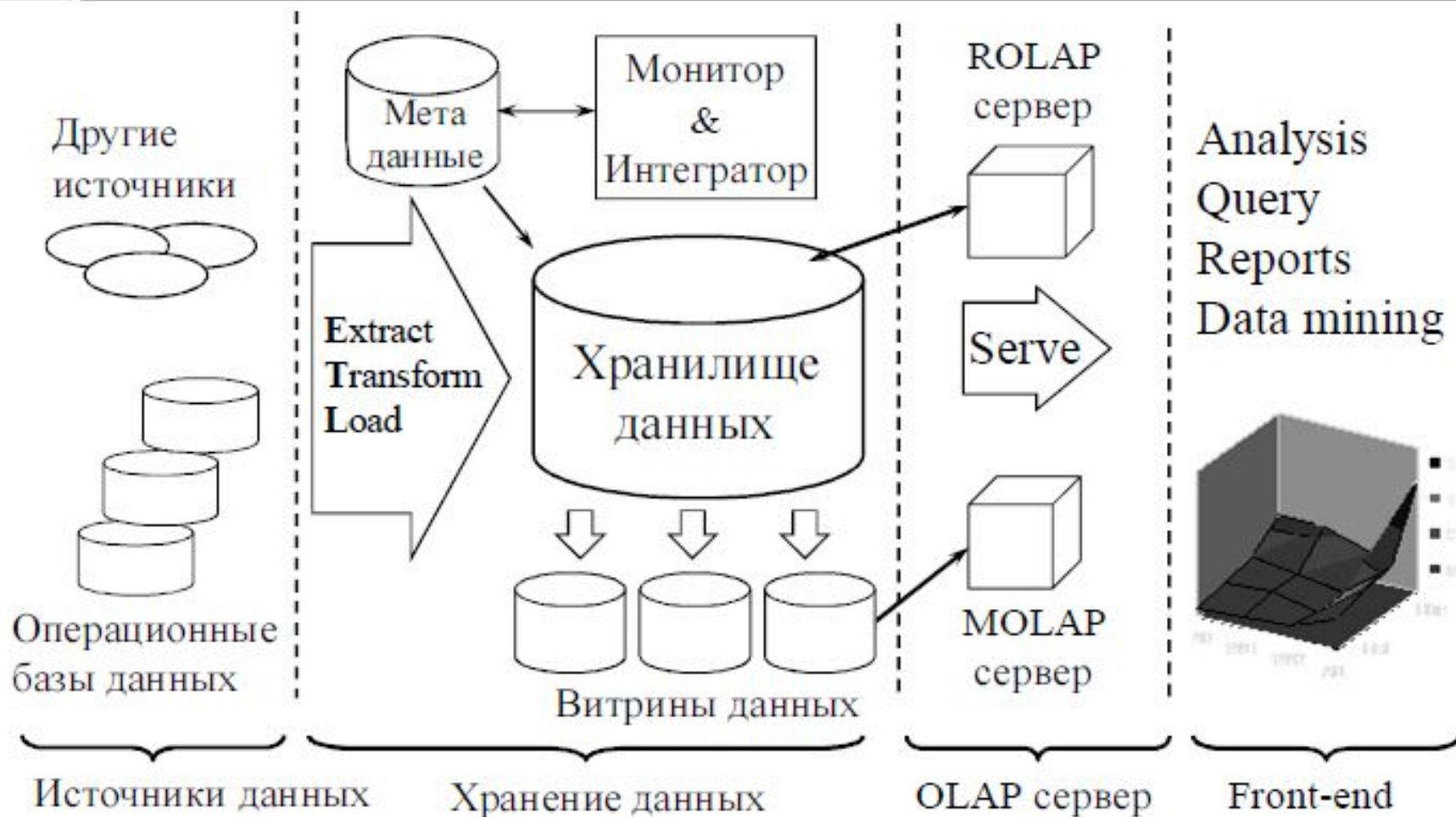
Проектирование хранилища

24

- Выбрать *бизнес-процесс* (поставки, продажи и др.)
- Выбрать *единицу (атомарный элемент данных)* бизнес-процесса
- Выбрать *измерения*
- Выбрать *меры*

Многоуровневое хранилище данных

25



Язык DMQL

28

- `define cube Продажи [Время, Товар, Филиал, Место]:`

 - `Выручка = sum(Сумма)`
 - `СрВыручка = avg(Сумма)`
 - `Вал = count(*)`

- `define dimension Время as (`
 - `КлючВремя, День, Неделя, Месяц, Квартал, Год)`
- `define dimension Товар as (`
 - `КлючТовар, НазТовар, Марка, Тип,`
 - `Поставщик(КлючПоставщик, ТипПоставщик))`
- `define dimension Филиал as (`
 - `КлючФилиал, НазФилиал, ТипФилиал)`
- `define dimension Место as (`
 - `КлючМесто, Улица,`
 - `Город(КлючГород, Округ, Страна))`