

# OLAP и другие технологии организации, обработки и хранения данных

ФИО преподавателя: Митина О.А.

e-mail: [alogmi@yandex.ru](mailto:alogmi@yandex.ru)

9

# Лекция

## Концепция хранилища данных. Понятие хранилища данных

# Условия обучения

- По итогам изучения дисциплины проводится экзамен
- В течение семестра необходимо выполнить все практические работы

# Основные положения концепции хранилищ данных

- интеграция и согласование данных из различных источников;
- разделение наборов данных, используемых системами обработки транзакций и системами поддержки принятия решений.

# Основные положения концепции хранилищ данных

**Предметно-ориентированный,  
интегрированный, неизменяемый и  
поддерживающий хронологию набор  
данных, предназначенный для обеспечения  
принятия управленческих решений.**

*Б. Инмон*

# Основные положения концепции хранилищ данных

**Хранилище** данных – разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая целостность, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов.

# Основные положения концепции хранилищ данных

**Отличие типичного ХД от обычных систем хранения данных:**

- цели использования;
- динамика изменения данных.

# Основные положения

## концепции хранилищ данных

- *высокая скорость получения данных из хранилищ;*
- *автоматическая поддержка внутренней непротиворечивости данных;*
- *возможность получения и сравнения срезов данных;*
- *наличие удобных средств для просмотра данных в хранилище;*
- *обеспечение целостности и достоверности хранящихся данных. (Р. Кимбалл)*



# Основные свойства хранилищ данных

- предметная ориентированность;
- интегрированность;
- неизменчивость;
- поддержка хронологии.

# Основные свойства хранилищ данных

**Под предметной ориентированностью** подразумевается, что ХД должно разрабатываться с учетом специфики конкретной предметной области, а не бизнес-приложений, с которыми его предполагается использовать.

# Основные свойства хранилищ данных

**Интегрированность** означает, что должна быть обеспечена возможность загрузки в ХД информации из источников, поддерживающих различные форматы данных и созданных в различных приложениях.

# Основные свойства хранилищ данных

Принцип **неизменчивости** предполагает, что, в отличие от обычных систем оперативной обработки данных, в ХД данные после загрузки не должны подвергаться каким-либо изменениям за исключением добавления новых данных.

# Основные свойства хранилищ данных

**Поддержка хронологии** означает соблюдение порядка следования записей, для чего в структуру ХД вводятся ключевые атрибуты **Дата** и **Время**.

# Детализированные и агрегированные данные

Данные в **детализированном** виде поступают непосредственно из первичных источников данных и соответствуют элементарным событиям, регистрируемым OLTP-системами.

# Детализированные и агрегированные данные

**Агрегирование** – процесс обобщения детализированных данных.

*Обобщенные данные называют агрегированными (агрегатами).*

# Метаданные (от греч. *meta* и лат. *data*)

**Метаданные** – высокоуровневые средства отражения информационной модели и описания структуры данных, используемые в ХД.



# Виды метаданных

Бизнес-метаданные

Технические  
метаданные

Операционные  
метаданные

Проектные метаданные

# Виды метаданных

## Бизнес-метаданные

представляют собой описание предметной области, для работы в которой создается ХД или витрина данных.

# Виды метаданных

**Технические метаданные** – это имена таблиц и названия столбцов в таблицах на физическом уровне разработки ХД.

# Виды метаданных

**Операционные метаданные  
необходимы для обеспечения  
функционирования хранилища.**

# Виды метаданных

**Проектные метаданные** – различная документация к хранилищу данных, регламенты его обслуживания, версионность и т. п.

# Модели хранилищ данных

- *реляционная;*
- *многомерная;*
- *гибридная;*
- *виртуальные.*

# Модели хранилищ данных

**Реляционная модель ХД** использует классическую реляционную модель баз данных.

*ROLAP – Relational OLAP*

# Модели хранилищ данных

**Многомерная модель ХД реализует многомерное представление данных на физическом уровне в виде многомерных кубов.**

*MOLAP – Multidimensional OLAP.*



# Модели хранилищ данных

**Гибридные ХД** сочетают в себе свойства как реляционной, так и многомерной моделей данных.

# Модели хранилищ данных

**Виртуальные ХД** – это набор баз данных, которые можно использовать совместно, чтобы пользователь мог получать доступ ко всем данным, как если бы они хранились в одном хранилище данных.

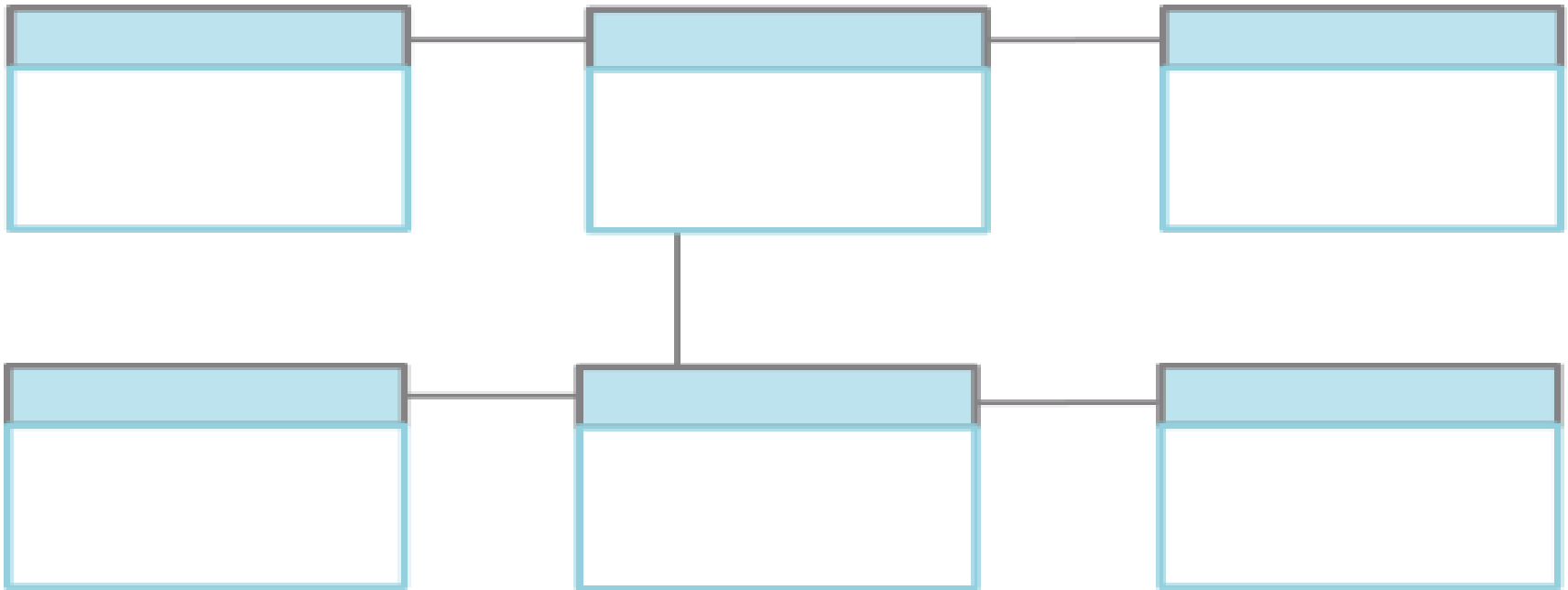
# Модели хранилищ данных

- одноплатформенные;
- кросс-платформенные.

# Многомерная модель данных

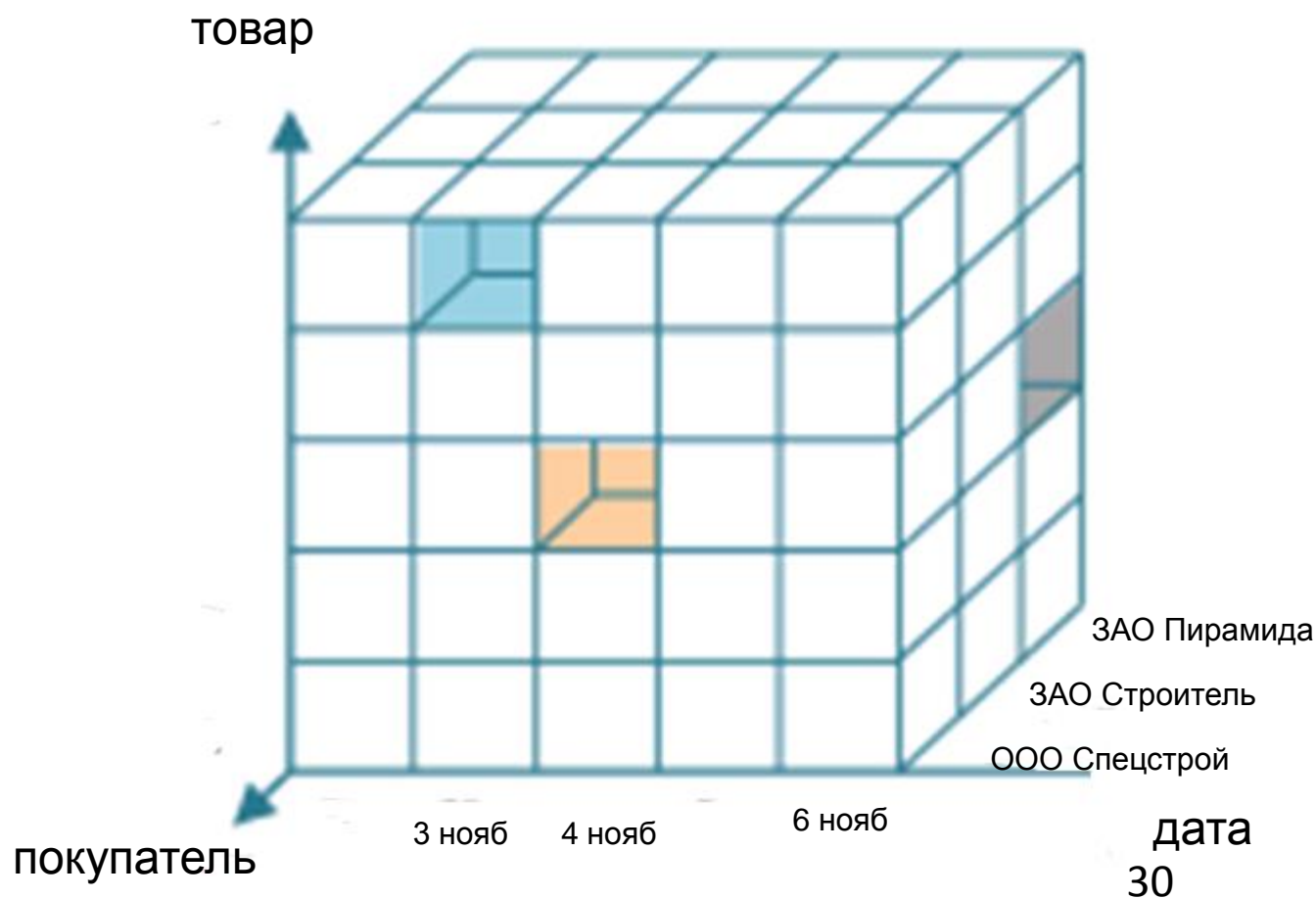
*OLAP-куб* (аббревиатура OLAP расшифровывается как *On-Line Analytical Processing* – оперативная аналитическая обработка).

# Концепция многомерного представления данных



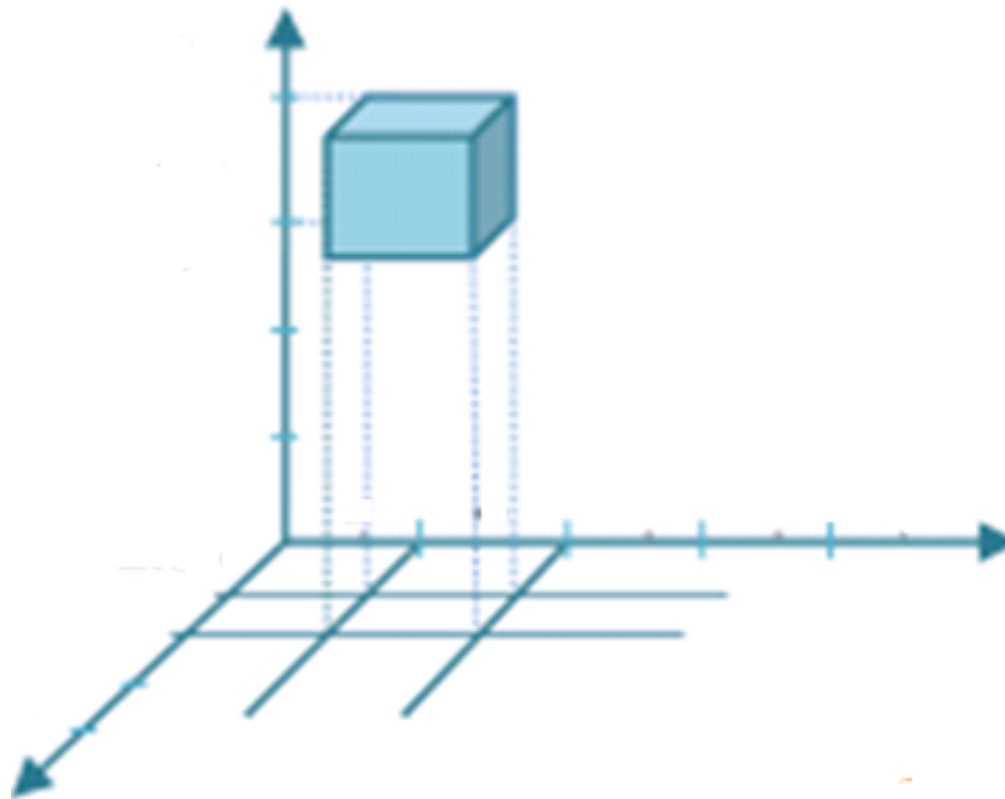
Реляционные модели баз данных с высокой степенью нормализации плохо подходят для хранения данных

# Структура многомерного куба



# Структура многомерного куба

товар



дата

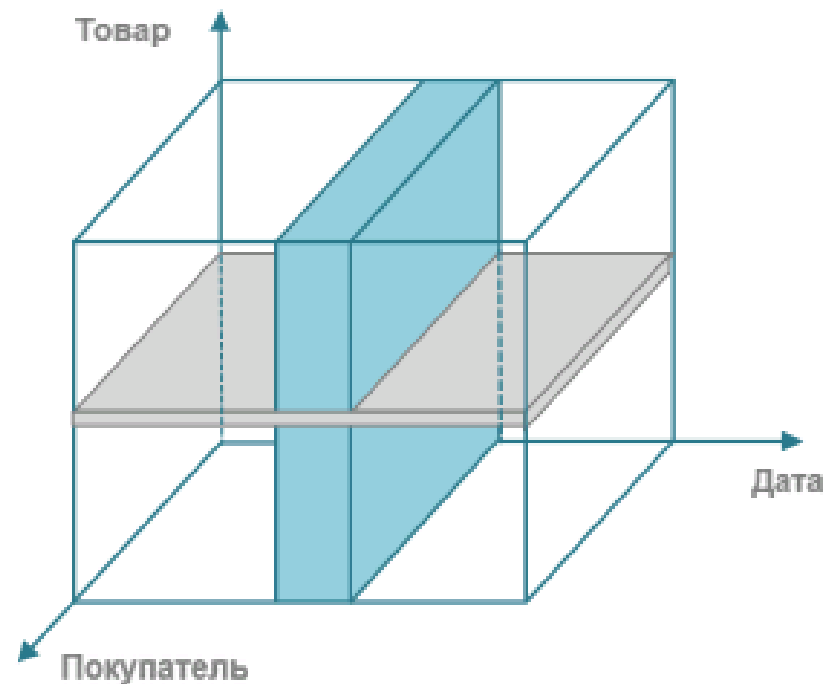
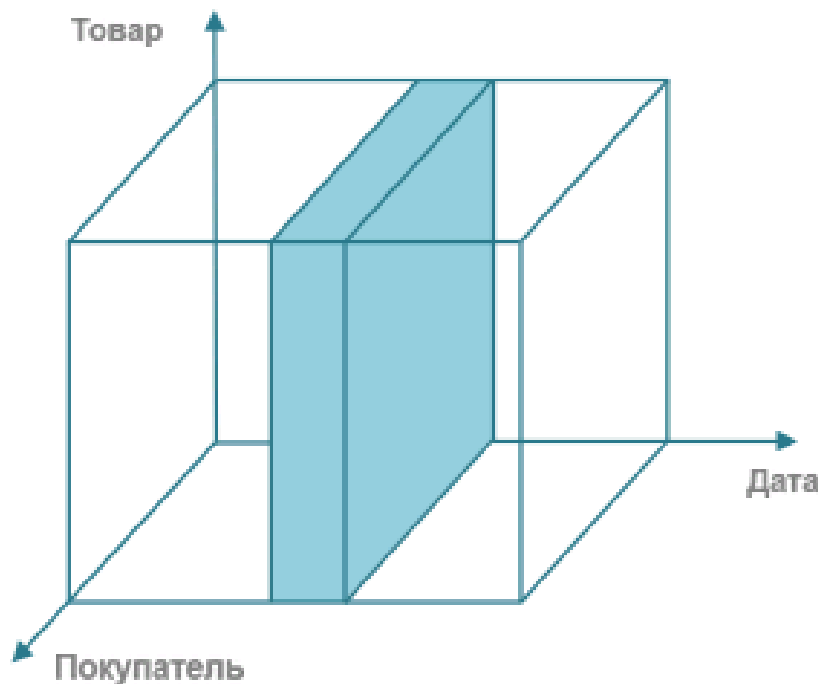
покупатель

# Работа с измерениями

- *сечение;*
- *транспонирование;*
- *сверстка;*
- *детализация.*



# Работа с измерениями



# Работа с измерениями

Группа	Товар	Σ Сумма
Инструмент	Дрель	2 450,00
	Отвертка	1 200,00
	Шпатель	780,00
	Электропила	7 600,00
Стройматериалы	Доска	7 400,00
	Керамзит	4 500,00
	Кирпич	22 000,00
	Цемент	12 000,00

# Агрегаты и нерегламентированные запросы

*Агрегатами называют агрегированные по определенным условиям исходные значения показателей.*

# Простой пример MOLAP

Дата	Товар	Покупатель	Сумма, руб.
02.11.2013	Плиты	ООО Спецстрой	120 600,00
02.11.2013	Плиты	ЗАО Строитель	100 500,00
03.11.2013	Бетон	ЗАО Пирамида	165 900,00
03.11.2013	Бетон	ООО Спецстрой	99 000,00
03.11.2013	Цемент	ЗАО Пирамида	12 000,00
03.11.2013	Кирпич	ЗАО Пирамида	15 850,00
04.11.2013	Кирпич	ООО Спецстрой	32 000,00
04.11.2013	Кирпич	ЗАО Пирамида	60 000,00
05.11.2013	Плиты	ЗАО Строитель	69 000,00
06.11.2013	Керамзит	ООО Спецстрой	45 930,00
06.11.2013	Плиты	ООО Спецстрой	56 000,00
06.11.2013	Цемент	ЗАО Строитель	51 100,00

# Простой пример MOLAP

Дата	Покупатель	Товар				
		Плиты	Бетон	Кирпич	Керамзит	Цемент
02.11.2013	ООО Спецстрой	120 600				
	ЗАО Строитель	100 500				
	ЗАО Пирамида					
03.11.2013	ООО Спецстрой		99 000			
	ЗАО Строитель					
	ЗАО Пирамида		1 65 900	15 850		12 000
04.11.2013	ООО Спецстрой			32 000		
	ЗАО Строитель					
	ЗАО Пирамида			60 000		
05.11.2013	ООО Спецстрой					
	ЗАО Строитель	69 000				
	ЗАО Пирамида					
06.11.2013	ООО Спецстрой	56 000			45 930	
	ЗАО Строитель					51 100
	ЗАО Пирамида			37		

# ROLAP – Реляционные хранилища

*Реляционная БД – это совокупность отношений, содержащих всю информацию, которая должна храниться в базе .*

*Э. Кодд*

# Схема «звезда» и «созвездие»

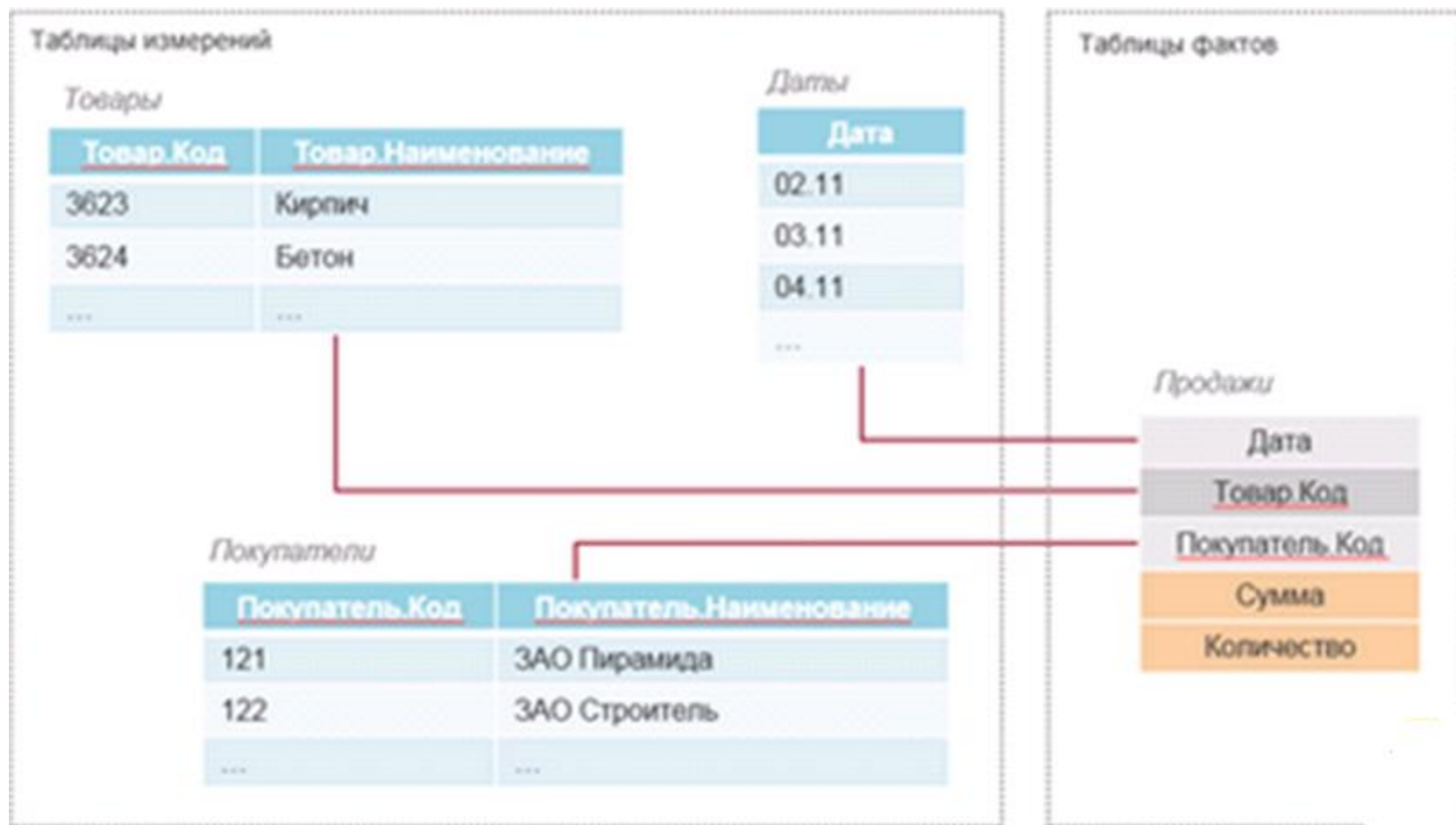


Схема «звезда»

# Схема «звезда» и «созвездие»

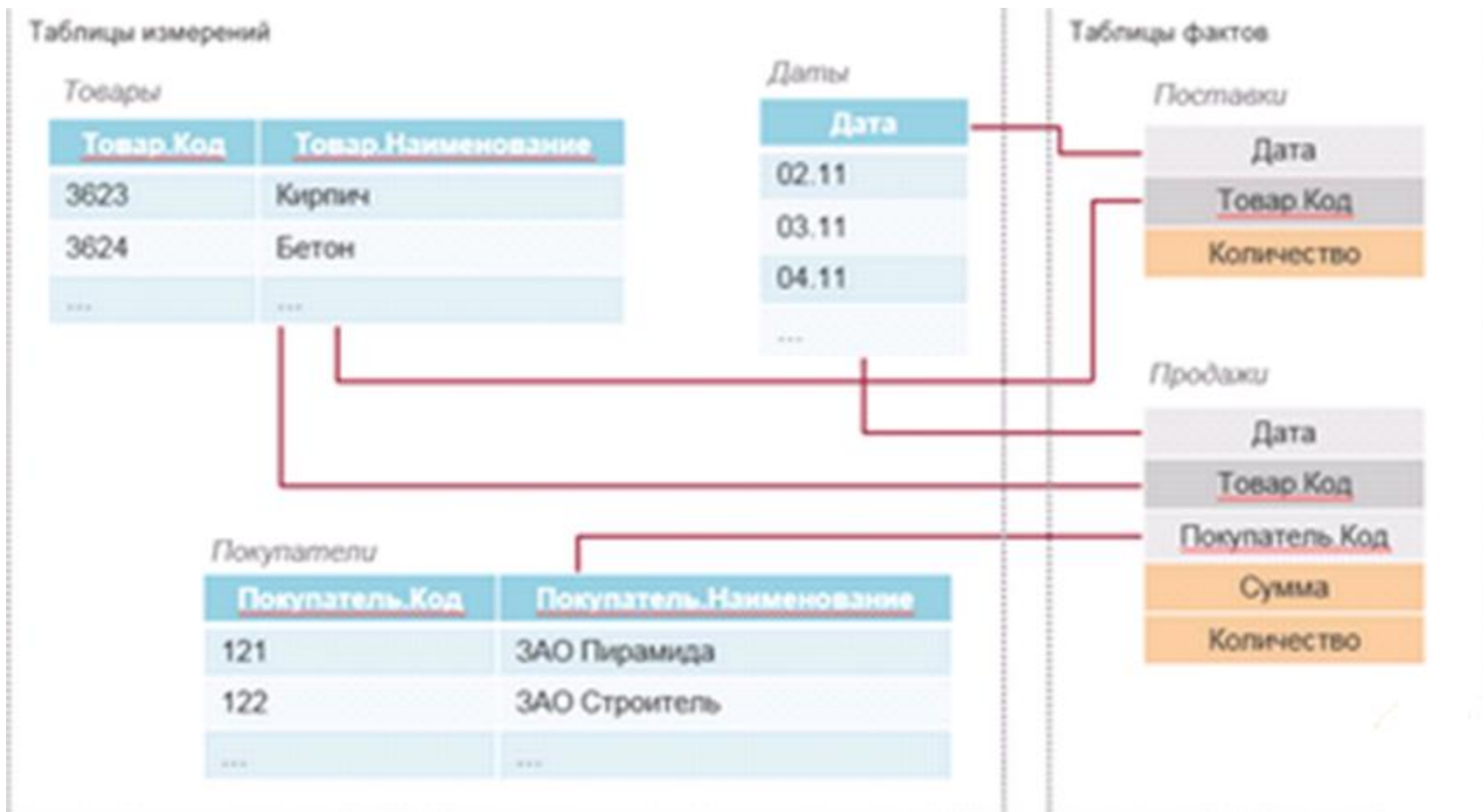


Схема «созвездие»



# Схема «снежинка»



Схема «снежинка»

# Сравнение схем

*Преимущества схемы «звезда»:*

- *простота и логическая прозрачность;*
- *более простая процедура пополнения измерений.*

# Сравнение схем

*Недостатки схемы «звезда»:*

- *медленная обработка измерений;*
- *высокая вероятность возникновения несоответствий в данных .*

## *Преимущества схемы «снежинка»:*

- намного ниже вероятность появления ошибок, несоответствия данных;*
- большая компактность представления данных.*

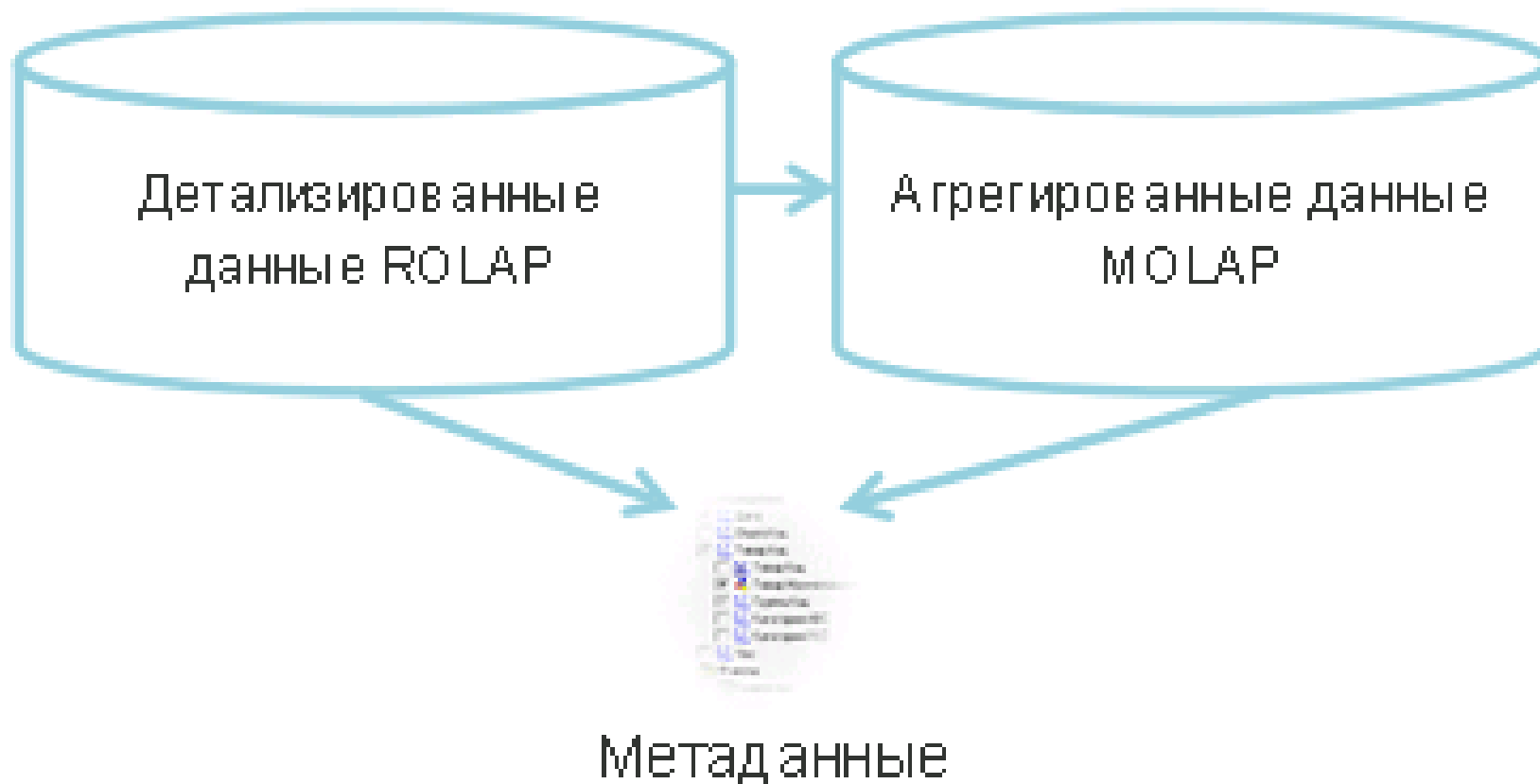
## *Недостатки схемы «снежинка»:*

- достаточно сложная для реализации и понимания структура данных;*
- усложненная процедура добавления значений измерений .*

# Сравнение MOLAP и ROLAP

- *производительность;*
- *гибкость.*

# HOIAR – гибридная модель



**HOIAR-система**

# Список литературы

- Тюрин Ю.Н. Анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров. – М.: МЦНМО, 2016. – 368 с.
- Мхитарян В.С. Анализ данных: учебник для академического бакалавриата / под ред. В.С. Мхитаряна. – М.: Изд. Юрайт, 2017 – 490 с.
- Хрусталёв Е.М. Агрегация данных в OLAP-кубах. [http:// www . olap . ru /](http://www.olap.ru/)



# Темы дисциплины

- 1 Анализ данных. Основные понятия и определения
- 2 Бизнес-аналитика. Основные понятия и определения
- 3 Концепция хранилища данных. Понятие хранилища данных
- 4 Многомерная модель данных
- 5-6 Интеграция данных и бизнес-аналитика
- 7-8 Интеграция данных
- 9 Хранилища данных
- 10 Процессы информативной корпоративной фабрики
- 11 Базовые архитектуры корпоративной информационной фабрики
- 12 Технология OLAP и ее особенности
- 13 Понятие OLAP-куба. Операции над OLAP-кубами
- 14 Аналитические платформы. Инструменты бизнес-аналитики
- 15-16 Большие данные. Наука о данных

# Спасибо за внимание!