

Модели данных

A5_Многомерные модели данных



Московский государственный технический университет
имени Н.Э. Баумана

Факультет ИБМ

Июль 2024 года

Москва

Артемьев Валерий Иванович © 2024

8. Модели реляционных витрин данных ROLAP

- Определение витрины данных
- Многомерная модель данных (гиперкуб)
- Основные элементы гиперкуба (факты, измерения, атрибуты)
- Простые и иерархические измерения (сбалансированные, неровные, несбалансированные, альтернативные иерархии)
- Таксономия многомерной модели данных
- Пример таксономии гиперкуба
- Схема «снежинка» (snowflake)
- Схемы «звезда» (star) и «созвездие» (constellation)

Актуализировать,
когда устоится

Витрина данных (Data Mart)

Множество тематических баз данных, содержащих информацию по отдельным аспектам деятельности организации, для обработки средствами бизнес-аналитики.

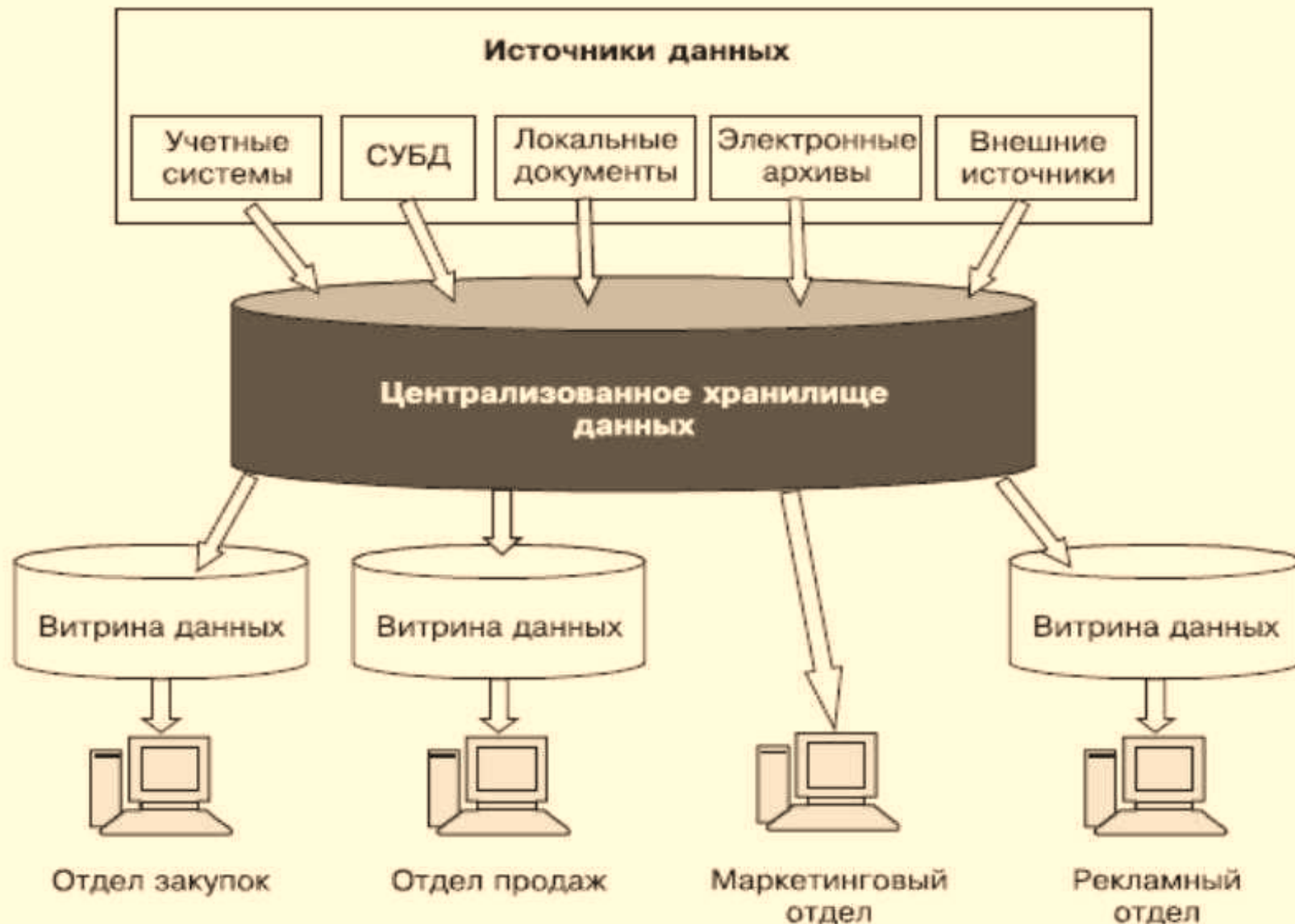
Основные цели выделения витрин:

- приближение данных к конечным пользователям,
- ограничение только необходимыми данными,
- повышение оперативности путём агрегирования данных,
- ориентация структуры данных на бизнес-аналитику.

Могут быть независимыми от хранилища данных или зависеть и быть частью хранилища данных.

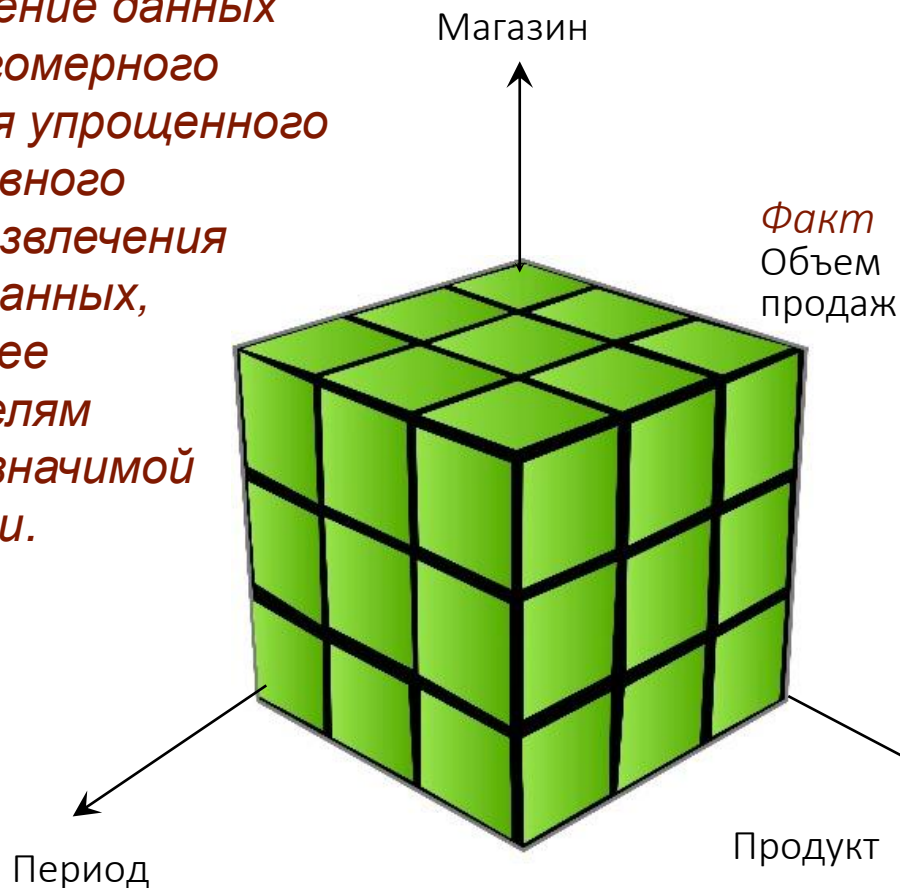
Часто используется многомерная абстракция для моделирования данных, реализации баз данных и обработки данных.

Место витрин данных в аналитических системах



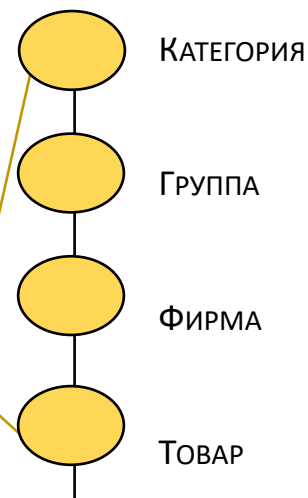
Многомерная модель данных (гиперкуб, факты, измерения)

Представление данных в виде многомерного массива для упрощенного и эффективного хранения, извлечения и анализа данных, облегчающее пользователям получение значимой информации.



Гиперкуб – многомерный массив данных, содержащий факты в нескольких измерениях.

Уровни



Измерения могут быть простыми, а также иерархическими

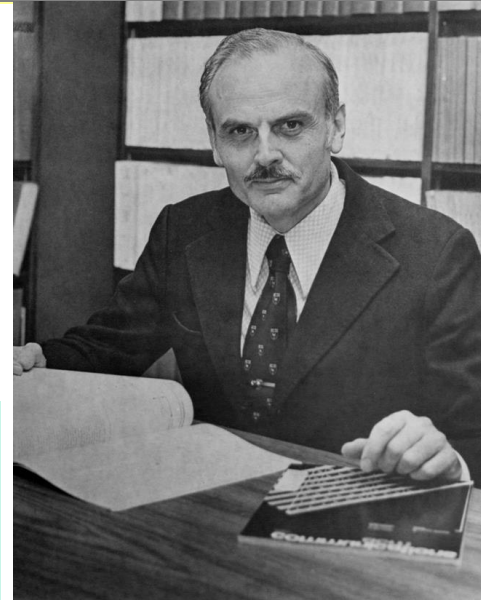
Обобщение реляционной модели от таблиц к массивам

Интерактивная аналитическая обработка (On Line Analytical Processing, OLAP)

Технологии анализа данных в различных разрезах и с разной степенью детальности, осуществляемого бизнес-пользователями в интерактивном режиме в терминах своей предметной области.



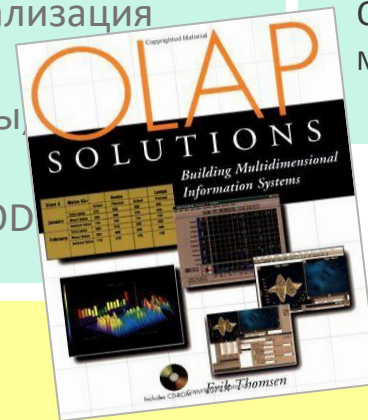
- Описательный и разведочный анализ
Многомерный анализ данных (детализация и укрупнение)
- Самообслуживание пользователей в бизнес-терминах
Привлечение опыта и интуиции пользователя
- «Ручная раскопка» данных (детализация и укрупнение)
Интерактивные запросы и отчёты
расчёты «на лету»
Инструменты NO CODE / LOW CODE



EDWARD CODD

IBM 1993

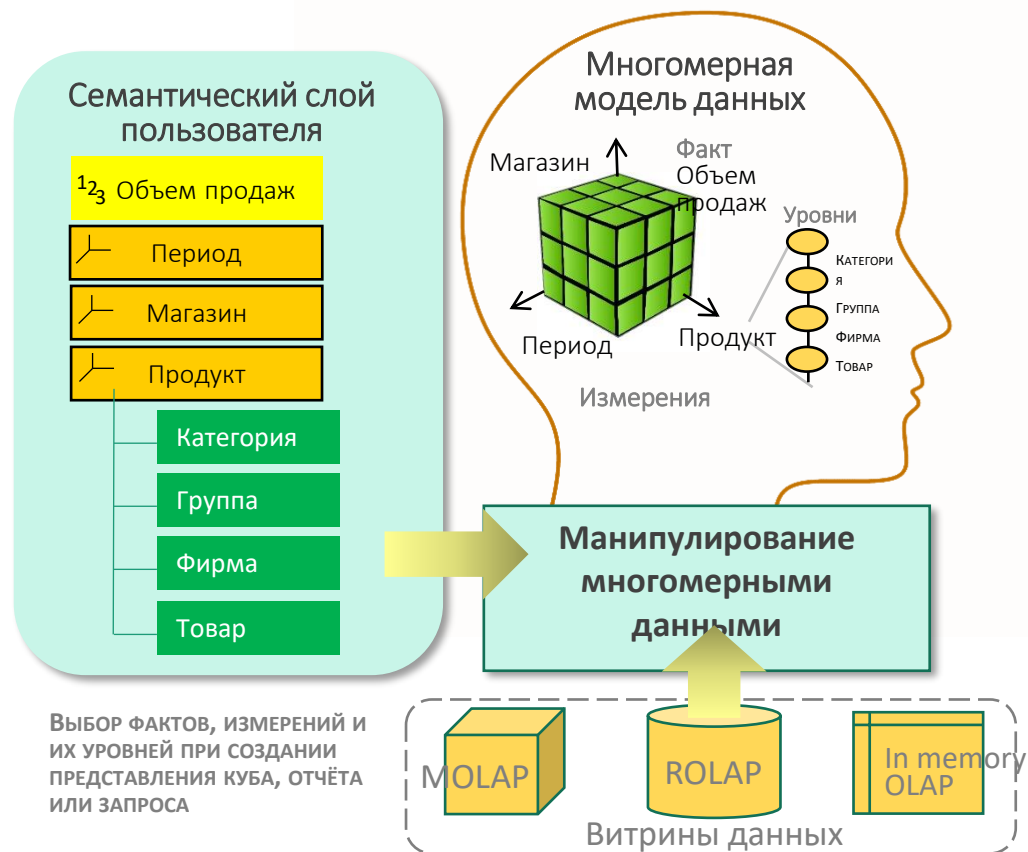
ОСНОВАТЕЛЬ РЕЛЯЦИОННОЙ
МОДЕЛИ ДАННЫХ



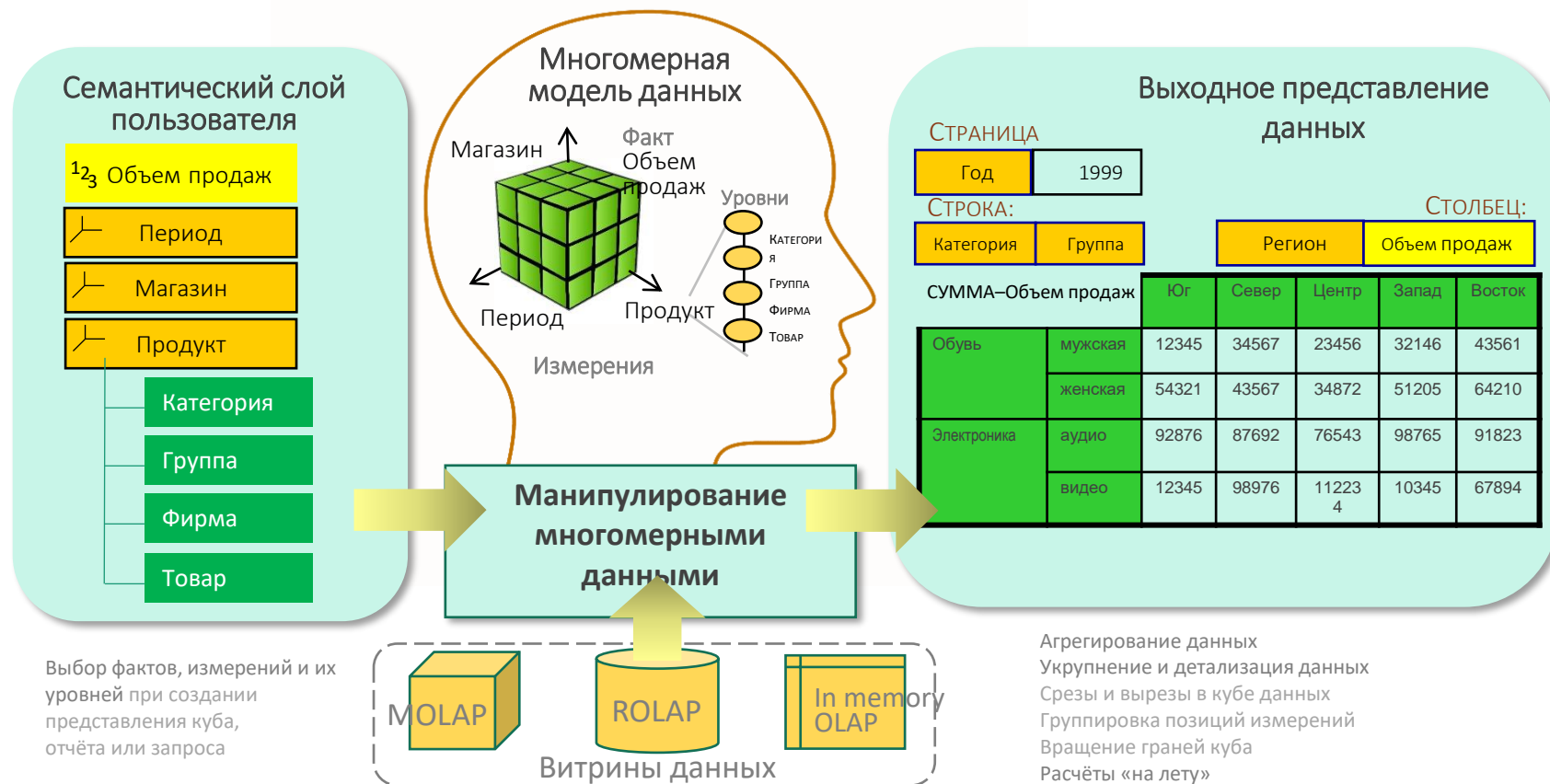
Многомерные свойства OLAP



Многомерные свойства OLAP

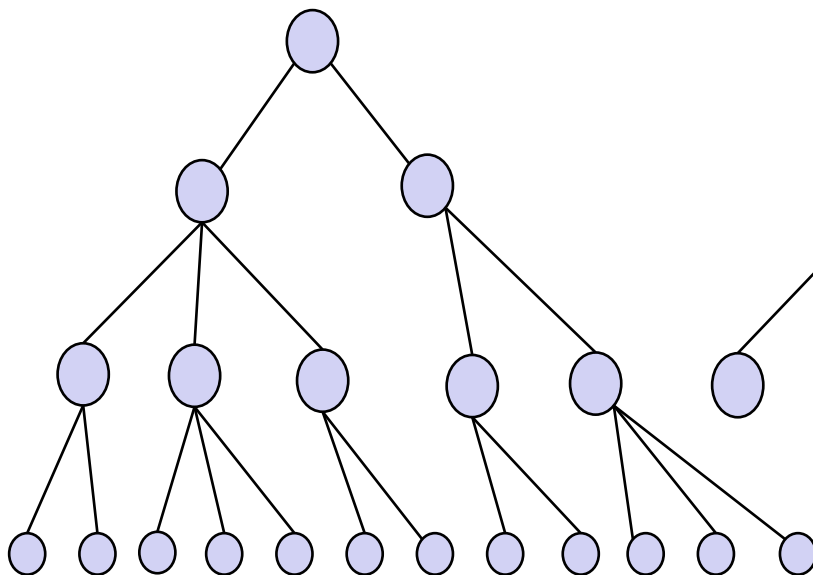


Многомерные свойства OLAP

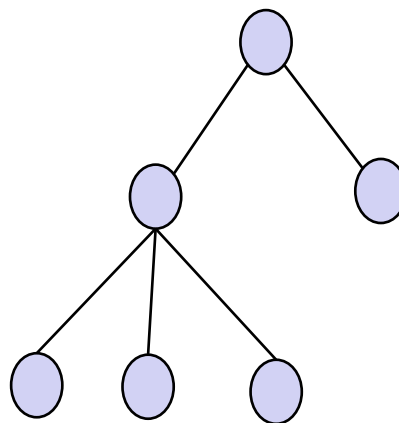


Виды иерархий измерений

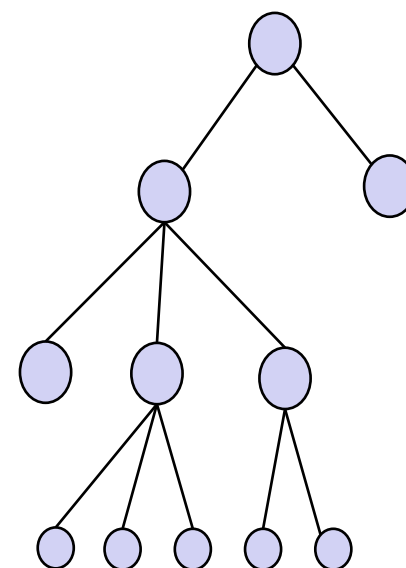
Сбалансированная



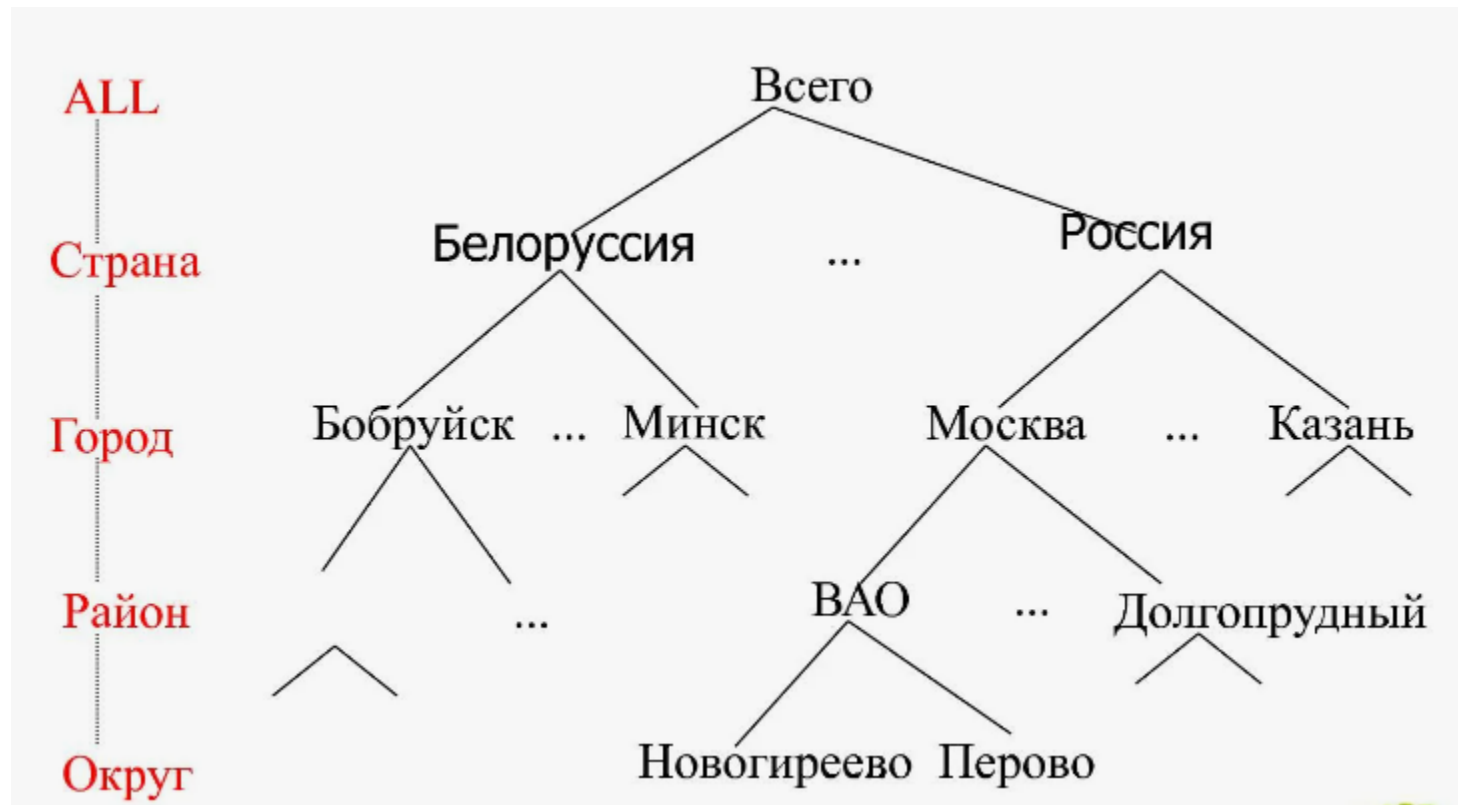
Неровная



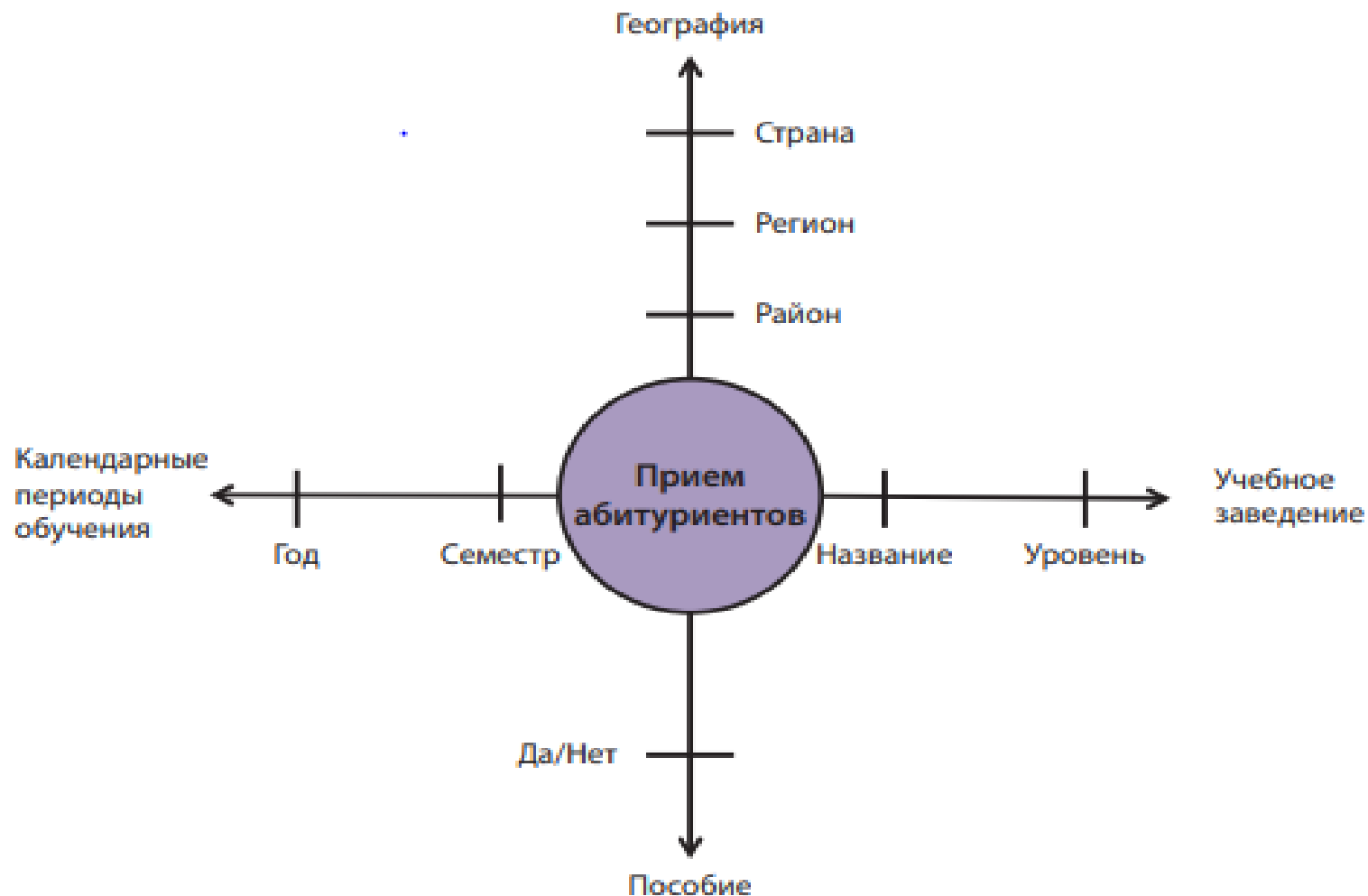
Несбалансированная



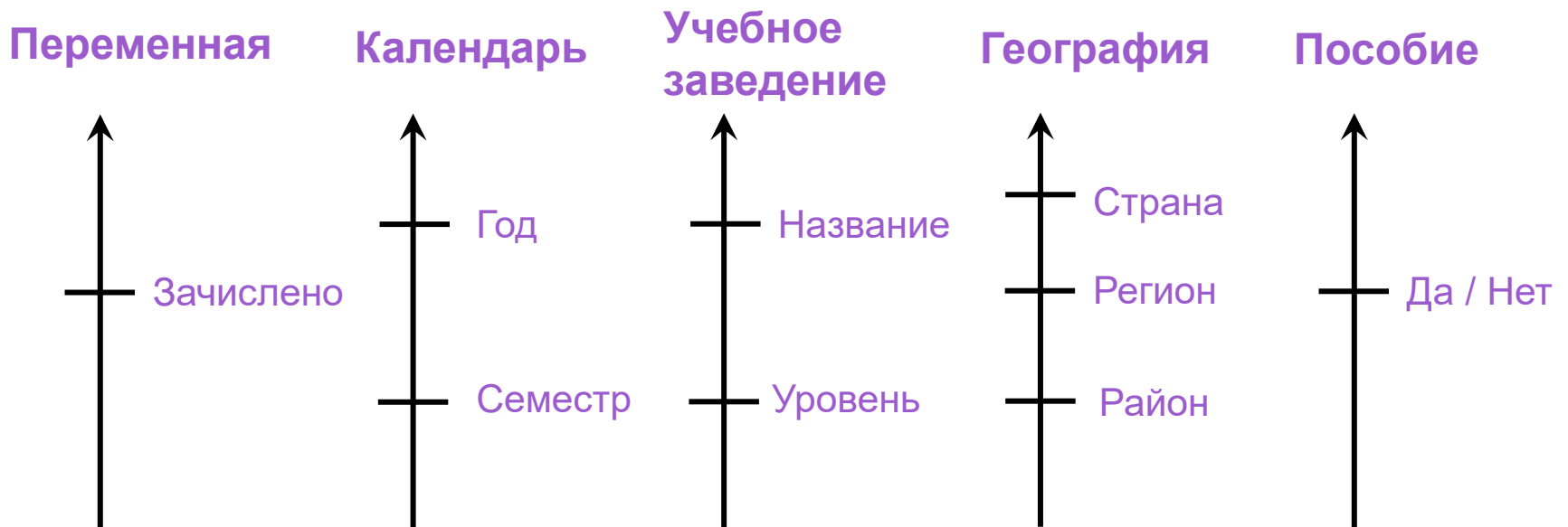
Пример иерархического измерения



Осевая нотация для представления многомерной модели данных



Другое представление многомерной модели данных



Таксономия многомерной модели данных (пример)

⊕ Приём абитуриентов (гиперкуб)

⊕ ФАКТЫ

⊖ Зачислено

⊕ ИЗМЕРЕНИЯ

⊕ Календарь

⊖ Год

⊖ Семестр

⊕ Учебное заведение

⊖ Название

⊖ Уровень

⊕ География

⊖ Страна

⊖ Регион

⊖ Район/

Реализация многомерных данных

MOLAP – многомерная база данных – многомерный массив на диске с прямо адресуемыми ячейками. Агрегаты и детали могут храниться в одном гиперкубе. Как правило, урезанный атрибутный состав.

ROLAP – реляционная база данных, имитирующая гиперкуб. Агрегаты хранятся отдельно от деталей. Широкий атрибутный состав.

In memory OLAP – многомерная база данных в основной памяти, обычно в сжатом виде с хранением по колонкам.

Структуры данных ROLAP

В реляционном представлении многомерные данные организованы в таблицы двух видов:

1. Таблицы фактов содержат количественные данные для дальнейшего анализа:

- *транзакционные факты* на основе отдельных событий;
- «*моментальные снимки*» (snapshot) на основе состояний объекта в определённые моменты времени;
- *факты, связанные с элементами документа*;
- *факты, связанные с событиями и состоянием объекта* без подробностей.

2. Таблицы измерений предоставляют контекстную и описательную информацию о данных в таблицах фактов. Часто представляется в виде *схемы снежинки* или *звезды*.

Структура таблицы фактов ROLAP

Таблица фактов содержит:

- факты одной степени детальности, соответствующие нижнему уровню иерархии;
- внешние ключи таблиц измерений,
- которые являются составным первичным ключом.

ПРОДАЖИ

(PK, FK) Календарь_Ид

(PK, FK) Магазин_Ид

(PK, FK) Продукт_Ид

Сумма

Количество

Агрегаты вычисляются «на лету» по первичным фактам или хранятся в отдельных таблицах фактов.

Структура таблицы измерений ROLAP

Таблица измерений содержит:

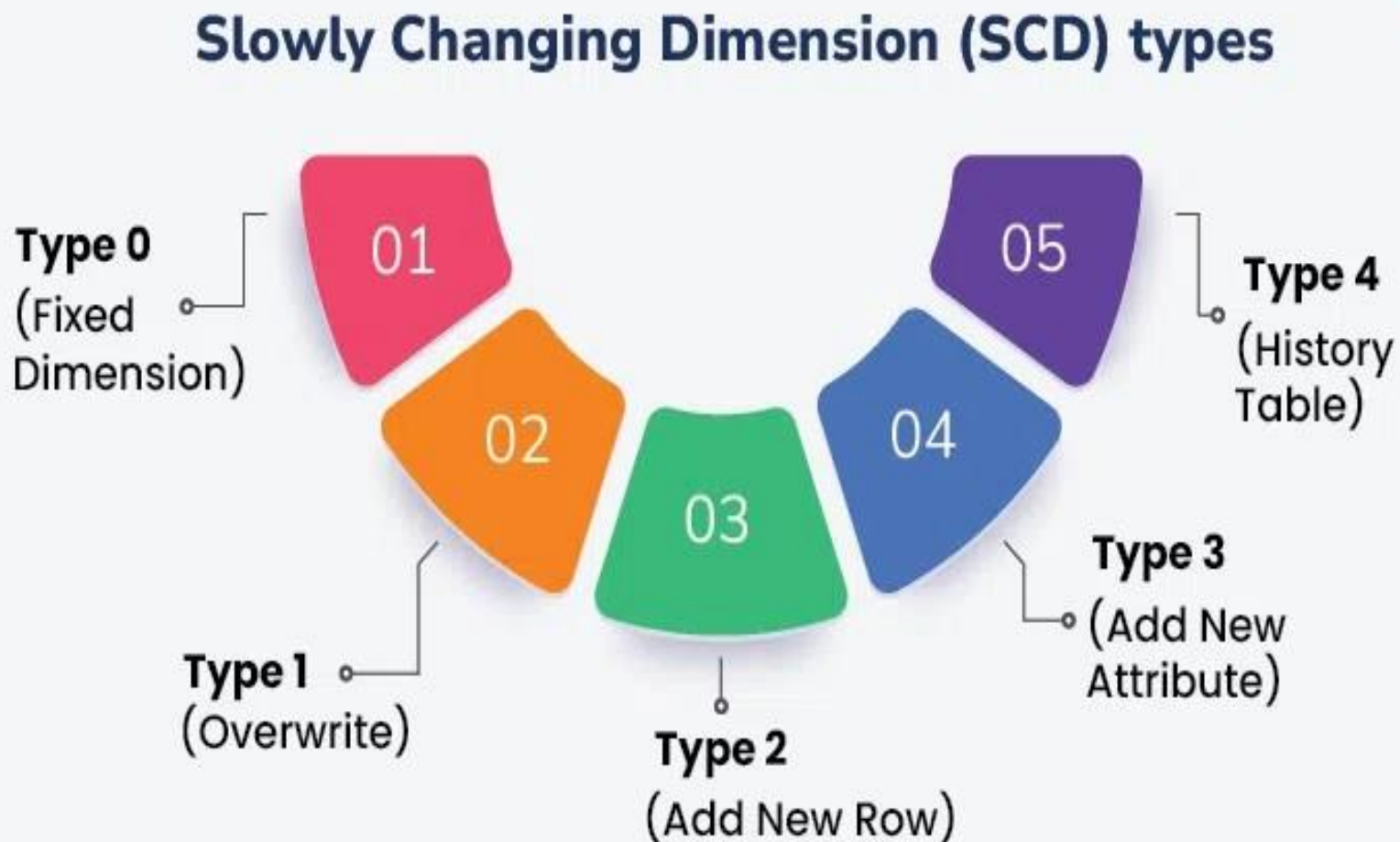
- по одной записи для каждого члена нижнего уровня;
- первичный ключ записи;
- характеристики нижнего уровня детальности;
- квалификаторы уровней агрегирования;
- все характеристики верхних уровней иерархии.

КАЛЕНДАРЬ

(PK) Месяц_Ид
Месяц_имя
Месяц_номер
Квартал_обозн
Квартал_номер
Год

Редко в таблице измерений хранится смесь разных степеней детальности, это затрудняет обработку и чревато ошибками.

Медленно меняющиеся размерности



Служебные атрибуты историчности

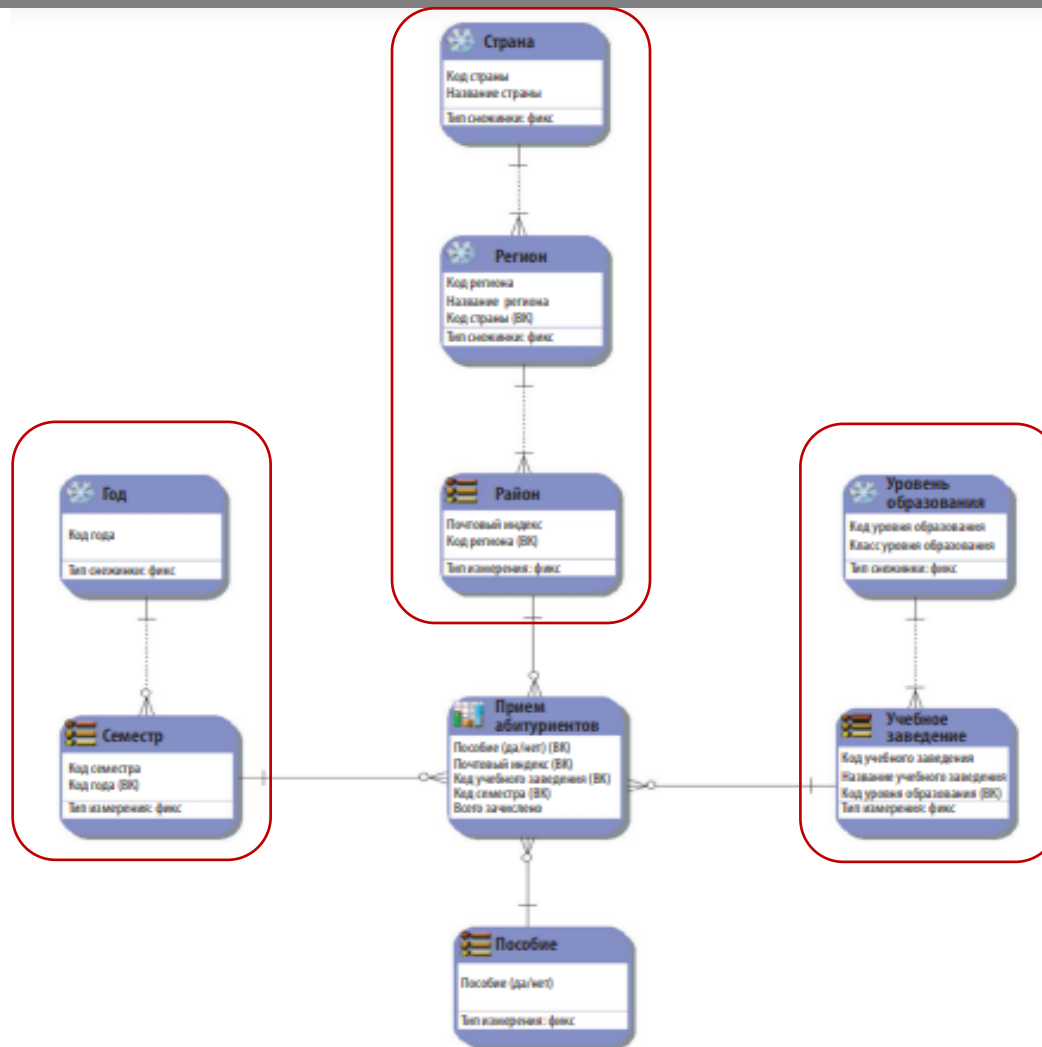
Для часто используемого способа SCD2 используют дополнительные атрибуты:

- Дата начала действия записи Effective_Date
- Дата конца действия записи End_Date
- Состояние актуальности Actual_Flag

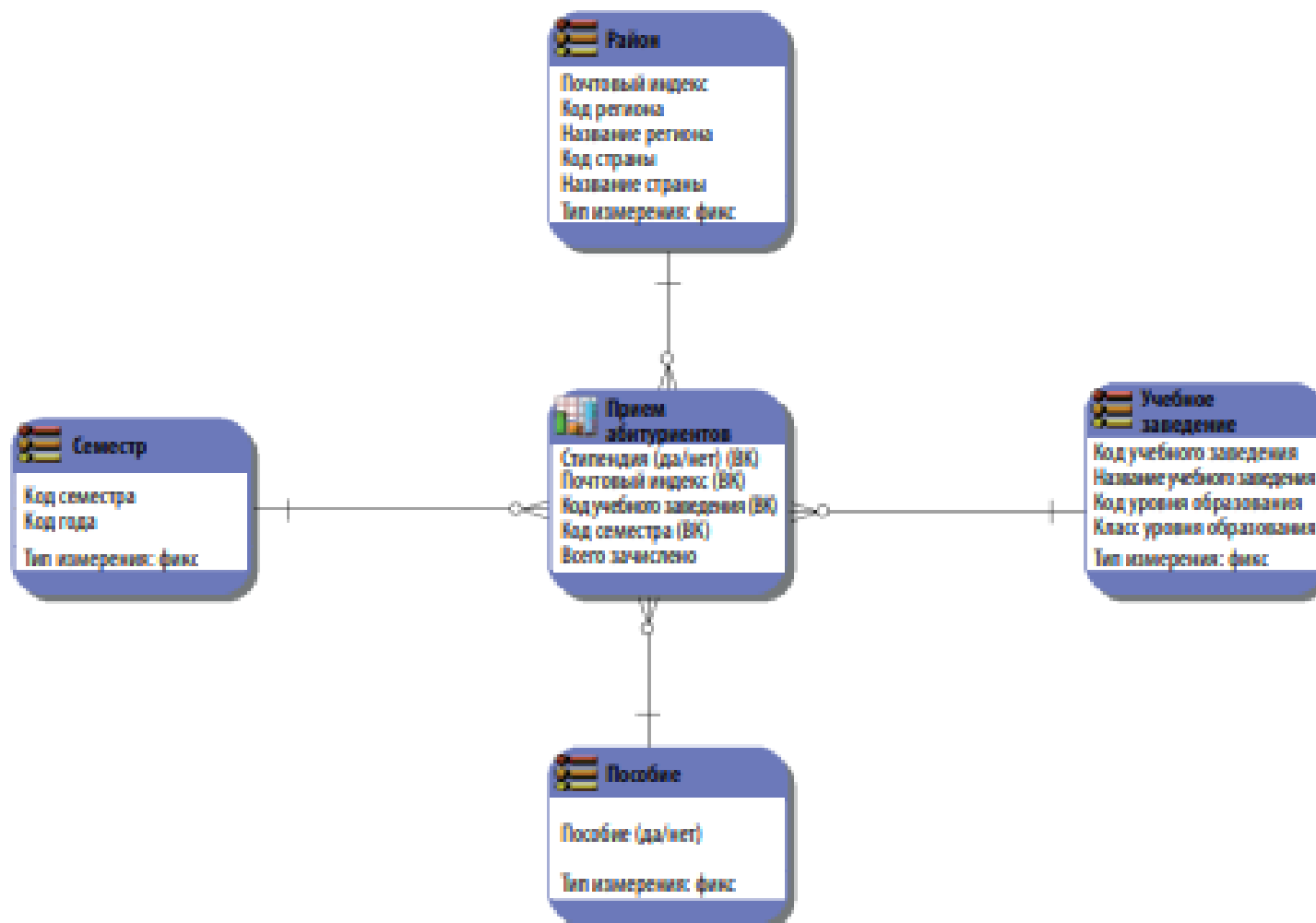
Для определения наличия и происхождения загруженных данных используют дополнительные атрибуты:

- Дата и время загрузки данных Load_Date_Time
- Код источника данных Source_Code

Схема «снежинка» (snowflake)



Схемы «звезда» (Star Schema)



Денормализация служит повышению скорости обработки

Терпения и удачи всем, кто связан
с моделированием данных

Спасибо за внимание!

Валерий Иванович Артемьев

Департамент статистики и управления данными
Банк России

Тел.: +7(495) 753-96-25

e-mail: avi@cbr.ru