

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ



Отличия информационных технологий КИС

- Изменяется масштаб, форма организации и средства управления БД - создаются хранилища и витрины данных.
- Используется компьютерная сеть Интранет, объединяющая в себе возможности локальных сетей, информационные технологии и сервис Интернет.
- Представлены два типа систем обработки данных, различающиеся составом и структурой данных, объемами хранимых и обрабатываемых данных, алгоритмами поиска и обработки данных и т.д.

Два типа систем обработки данных в КИС

- Система OLTP (On-Line transaction processing) — транзакционного типа.
- Система OLAP (On-Line Analytic Processing) — аналитического типа.

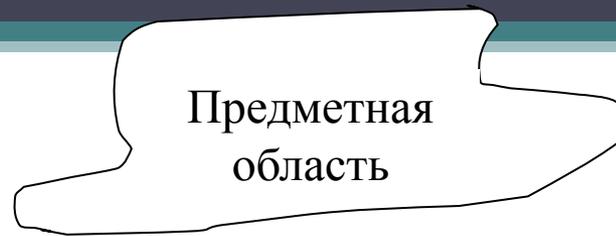
Два типа систем обработки данных в КИС

Показатель	OLTP	OLAP
Частота и объем обновляемых данных	Большая частота обновлений, относительно малый объем обновляемых данных	Низкая частота обновления и большой объем обновляемых данных
Горизонт хранения данных	Один календарный год	Неограниченный
Степень агрегирования данных	Первичные учетные данные	Промежуточные итоги. Сводные итоги. Консолидированные данные
Характер обработки данных	Строго регламентированные запросы и отчеты	Произвольные запросы, выборки, своды
Цели создания	Учет первичных данных. Оптимизация хранения данных. Оптимизация эксплуатационных характеристик приложений	Анализ и моделирование данных. Формирование прогнозов. Формализация знаний о системе управления

Информационные базы КИС

- базы данных (Data Base - DB, БД),
- хранилища данных (Data Warehouse - DW или ХД),
- базы знаний (Knowledge Base - KB, БЗ).

Базы данных



Уровень объектов ПрО

Логический уровень

Физический уровень

Концептуальная схема - абстрагированное описание предметной области с фиксированной (логической) точки зрения.

Логическое (концептуальное) проектирование - систематизация понятий и связей предметной области.

Модель данных - совокупность функциональных характеристик объектов и особенностей представления информации, используемая при абстрагировании.

Внутренняя схема - отображение концептуальной схемы на физический уровень.

Модель данных – модель логического уровня проектирования БД.

□ *иерархическая модель данных*

(hierarchical data model);

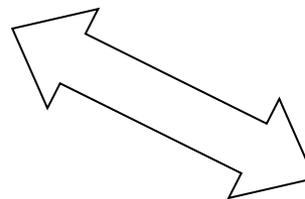
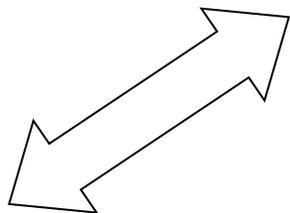
□ *сетевая модель данных* (network data model);

□ *реляционная модель данных* (relational data model).

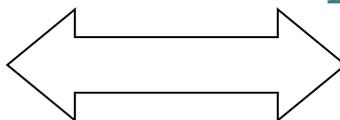
Модель данных

Структура данных

(Под **структурой данных** в общем случае понимают множество элементов данных и множество связей между ними.)



**Манипулирование
данными**



**Целостность
данных**

(Целостность данных означает систему правил, используемых для поддержания связей между записями в связанных таблицах, а также для обеспечения защиты от случайного удаления или изменения связанных данных.)

Реляционная модель данных

Структура

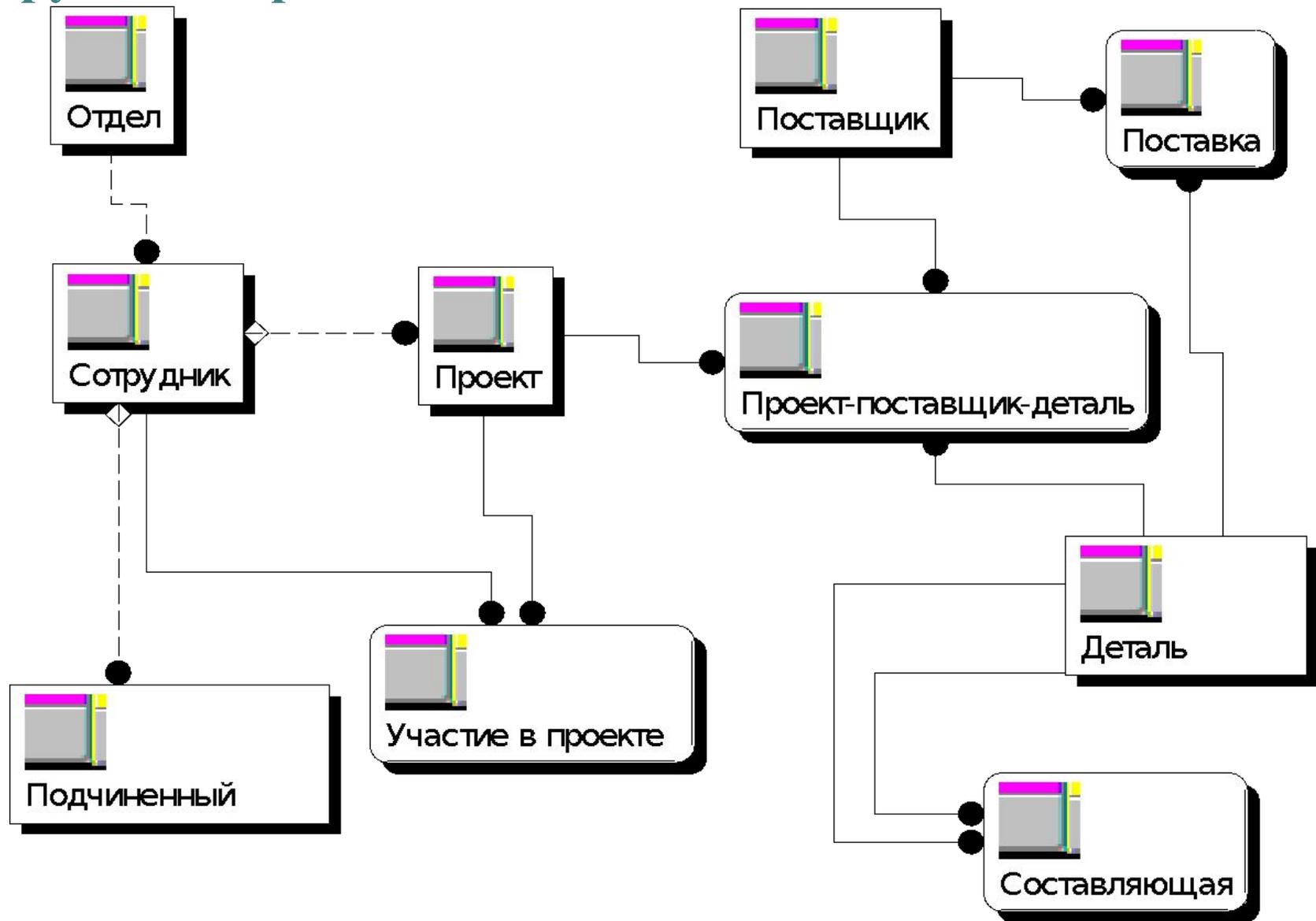
<i>Каф</i>	<i>Телефон</i>	<i>Корпус</i>	<i>№ ком</i>
22	25-15	А	322
23	38-42	В	221
24	99-18	Б	117

Кадровый состав

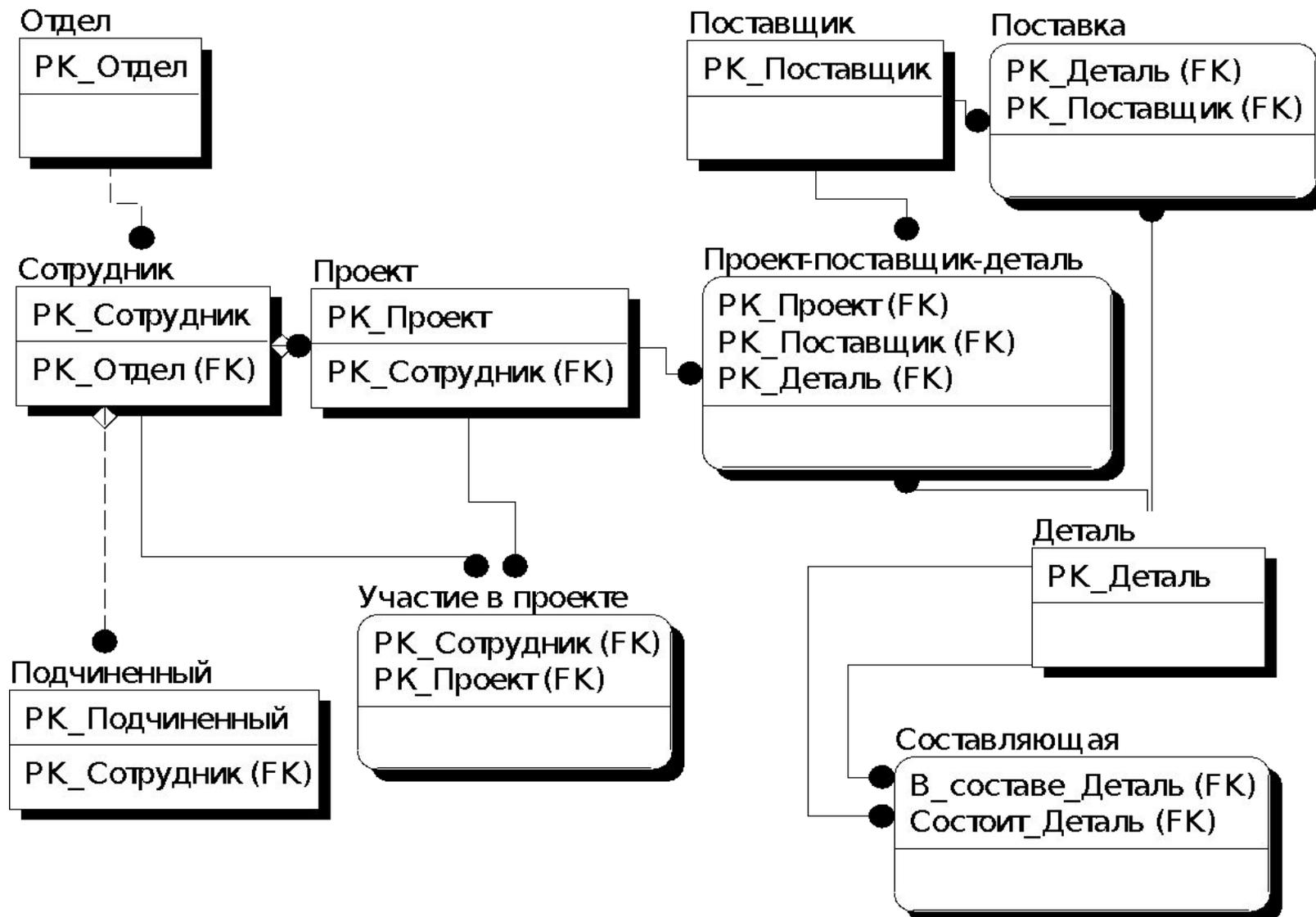
<i>Таб.№</i>	<i>ФИО</i>	<i>Должность</i>	<i>Каф</i>
121	Иванов И.И.	Зав. каф.	22
231	Сидоров С.С.	Проф.	22
123	Гиацинтова Г.Г.	Проф.	23
432	Цветкова С.С .	Доцент	23
465	Козлов К.К.	Доцент	24
675	Петров П.П.	Ст.преп.	24
782	Лютикова Л.Л.	Ассистент	22

Реляционная модель для ПрО

«сотрудники-проекты-детали-поставщики»



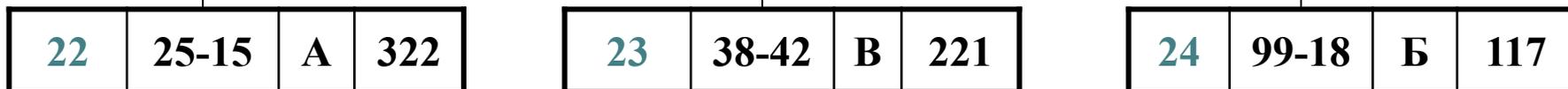
Реляционная модель (ключевые группы)



Иерархическая модель данных

ВУЗ

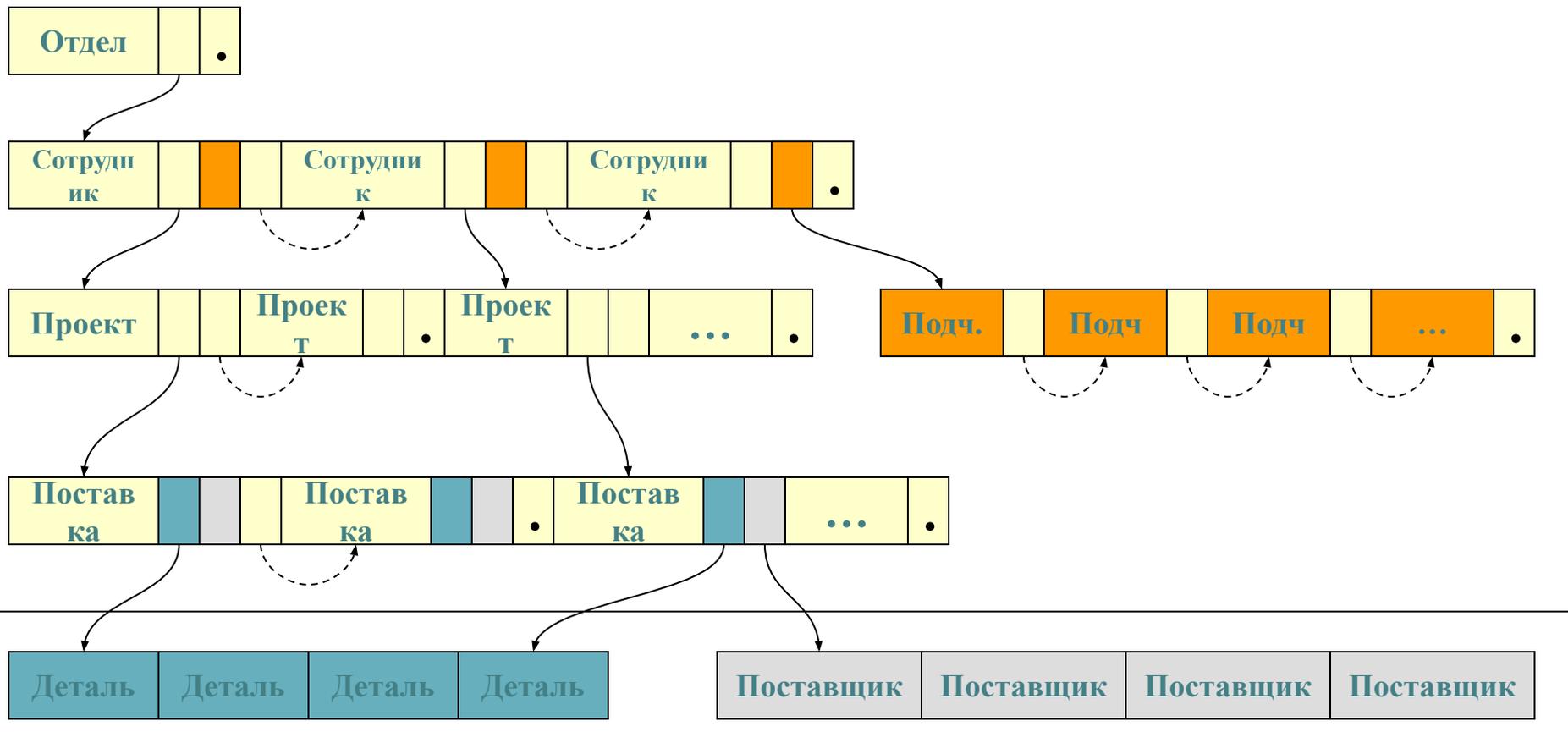
Структура



Кадровый состав

121	Иванов И.И.	Зав. каф.
231	Сидоров С.С.	Проф.
782	Лютикова Л.Л.	Ассистент
123	Гиацинтова Г.Г.	Проф.
432	Цветкова С.С .	Доцент
465	Козлов К.К.	Доцент
675	Петров П.П.	Ст.преп.

Иерархическая модель для ПрО «сотрудники-проекты-детали-поставщики»



Сетевая модель данных

Структура

22	25-15	А	322
----	-------	---	-----

23	38-42	В	221
----	-------	---	-----

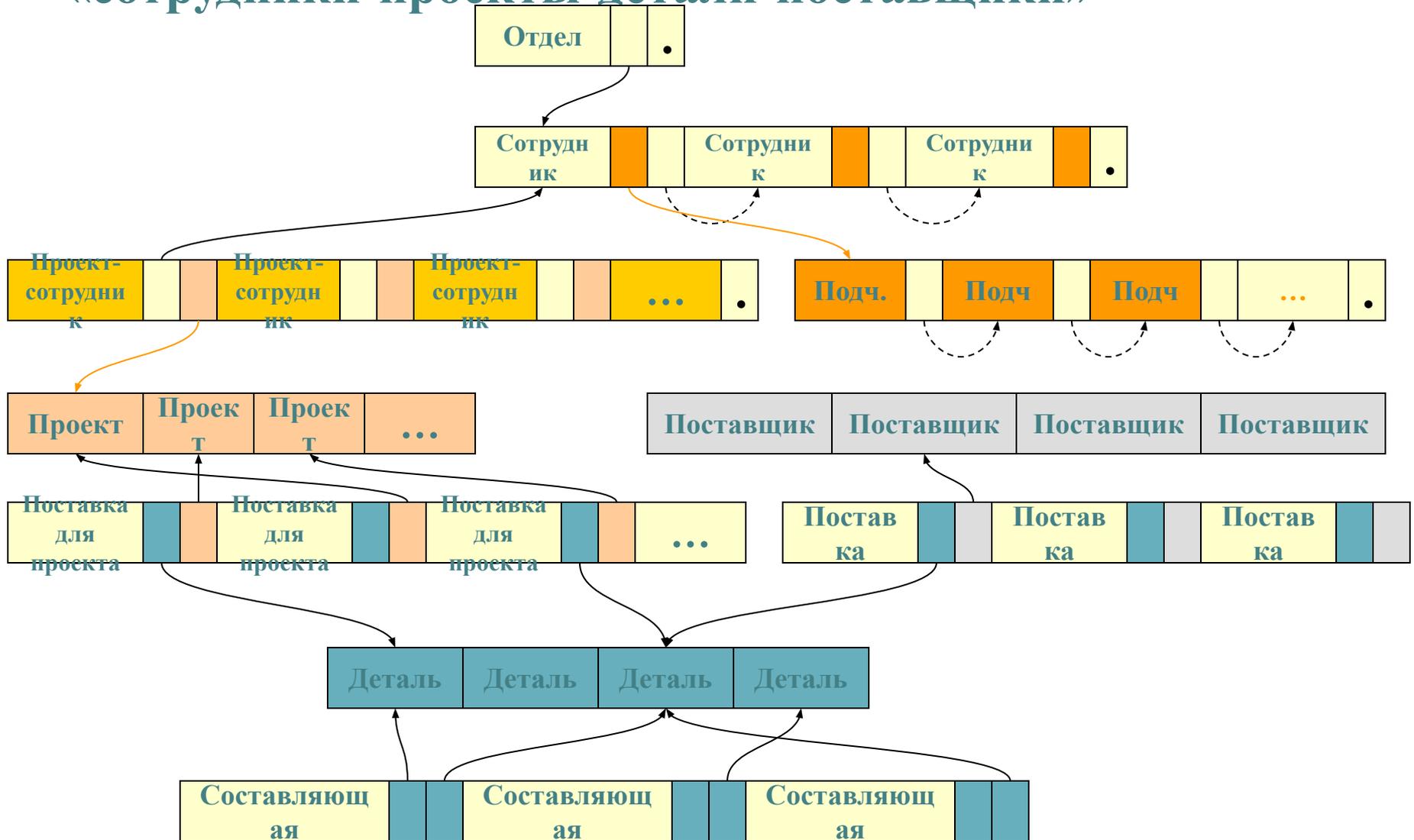
24	99-18	Б	117
----	-------	---	-----

Кадровый состав

121	Иванов И.И.	Зав. каф.
231	Сидоров С.С.	Проф.
782	Лютикова Л.Л.	Ассистент
123	Гиацинтова Г.Г.	Проф.
432	Цветкова С.С.	Доцент
465	Козлов К.К.	Доцент
675	Петров П.П.	Ст. преп.

Сетевая модель для ПРО

«сотрудники-проекты-детали-поставщики»



Преимущества и недостатки моделей

	<i>Иерархическая</i>	<i>Сетевая</i>	<i>Реляционная</i>
Связи между данными	Ссылочный аппарат	Ссылочный аппарат	Аппарат внешних ключей
Связи многие-ко-многим	С дублированием данных	Без дублирования данных	Без дублирования данных
Знание физической организации БД	Требуется	Требуется	Не требуется
Подход к обработке данных	Навигационный	Навигационный	Декларативный

Тип товара	Измерение	Наименование товара	Поставщик	Кол-во	Цена Руб.
Молоко	Бут.	Домик в деревне	Поставщик 1	30	45.00
		Простоквашино	Поставщик 2	10	50.00
			Поставщик 3	10	51.00
		Ромашкино	Поставщик 4	15	47.00
Сыр	кг	Пармезан	Поставщик 2	20	300.
			Поставщик 3	15	400
		Российский	Поставщик 4	10	250
Хлеб	шт	Нарезной батон	Поставщик 1	100	30
		Бородинский	Поставщик 2	50	35
		Измайловский	Поставщик 4	80	36
Овощи	кг	Картофель	Поставщик 2	100	20
			Поставщик 5	150	15

Тип товара	Измерение	Наименование товара	Поставщик	Кол-во	Цена Руб.
Молоко	Бут.	Домик в деревне	Поставщик 1	30	45.00
Молоко	Бут.	Простоквашино	Поставщик 2	10	50.00
Молоко	Бут.	Простоквашино	Поставщик 3	10	51.00
Молоко	Бут.	Ромашкино	Поставщик 4	15	47.00
Сыр	кг	Пармезан	Поставщик 2	20	300.
Сыр	кг	Пармезан	Поставщик 3	15	400
Сыр	кг	Российский	Поставщик 4	10	250
Хлеб	шт	Нарезной батон	Поставщик 1	100	30
Хлеб	шт	Бородинский	Поставщик 2	50	35
Хлеб	шт	Измайловский	Поставщик 4	80	36
Овощи	кг	Картофель	Поставщик 2	100	20
Овощи	кг	Картофель	Поставщик 5	150	15

Нормализация -

это процесс последовательной замены таблицы ее полными декомпозициями до тех пор, пока все они не будут находиться в 5НФ.

Процедура нормализации:

в таблице единственными функциональными зависимостями должны быть зависимости вида

$A \rightarrow K$

K - первичный ключ,

A - некоторый атрибут

Пример:

ХРАНЕНИЕ (ФИРМА, СКЛАД, ОБЪЕМ)

Содержит информацию о фирмах, получающих товары со складов, и объемах этих складов.

Каждая фирма может получать товар только с одного склада.

Функциональные зависимости (транзитивная):

- *фирма -> склад*
- *склад -> объем*

При этом возникают аномалии:

- если в данный момент ни одна фирма не получает товар со склада, то в базу данных нельзя ввести данные о его объеме (т.к. не определен ключевой атрибут)
- если объем склада изменяется, необходим просмотр всего отношения и изменение кортежей для всех фирм, связанных с данным складом.

Для устранения этих аномалий необходимо декомпозировать исходное отношение на два:

ХРАНЕНИЕ (ФИРМА, СКЛАД)

ОБЪЕМ_СКЛАДА (СКЛАД, ОБЪЕМ)

Пример:

ПОСТАВКИ (N_ПОСТАВЩИКА, ТОВАР, ЦЕНА, КОЛИЧЕСТВО)

Поставщик может поставлять различные товары. Один и тот же товар может поставляться разными поставщиками.

Все поставщики поставляют товар по одной и той же цене.

Ключ отношения: "*N_поставщика + товар*".

Функциональная зависимость

- *N_поставщика, товар -> количество*
- *товар -> цена*

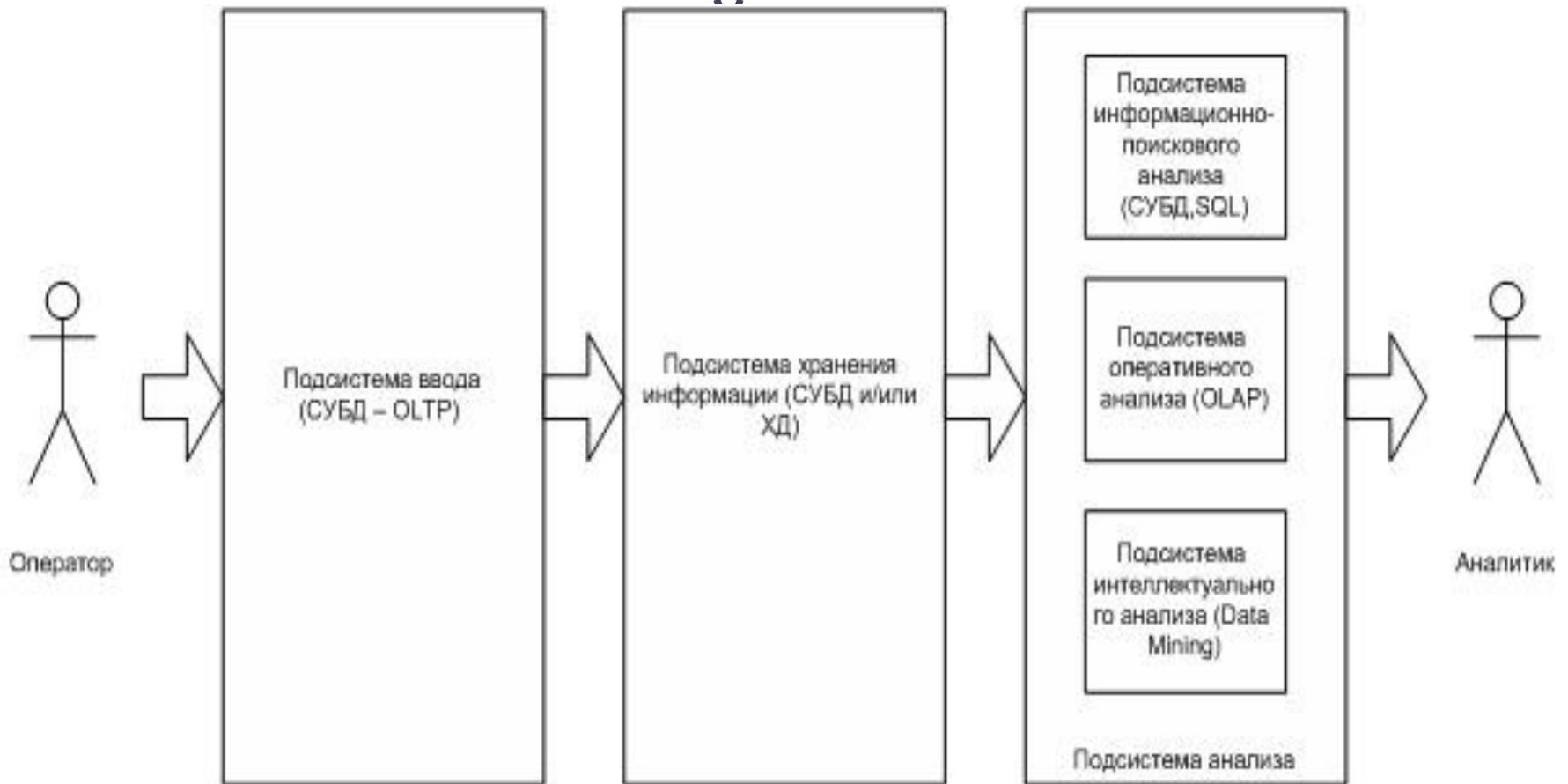
Следующее разложение дает отношения во 2НФ:

ПОСТАВКИ (N_ПОСТАВЩИКА, ТОВАР, КОЛИЧЕСТВО)

ЦЕНА_ТОВАРА (ТОВАР, ЦЕНА)

Хранилища данных

Архитектура системы поддержки



Хранилище данных

Технология ХД предназначена для хранения и анализа больших объемов данных с целью дальнейшего обнаружения в них скрытых закономерностей.

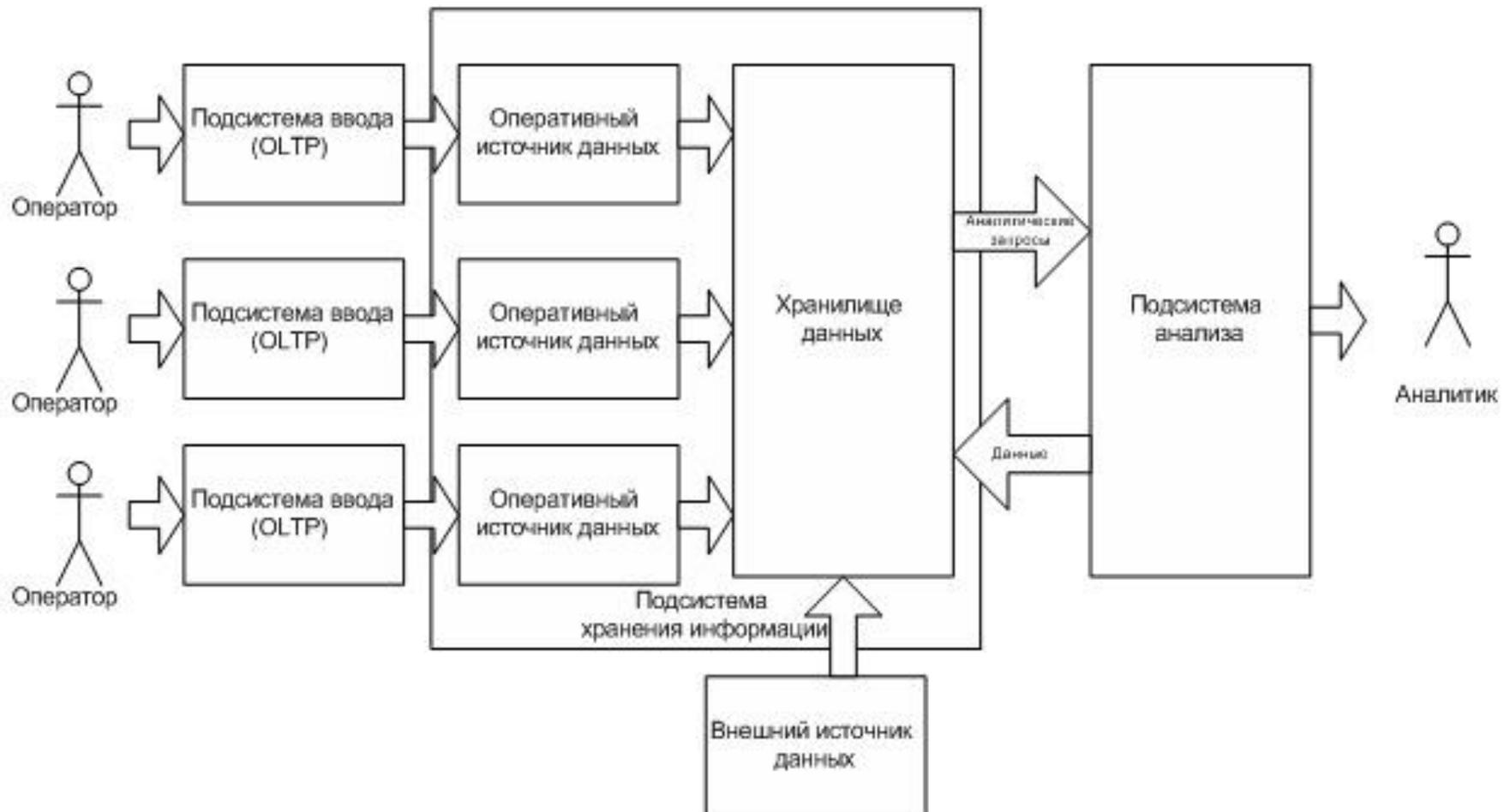
Data Mining изучает процесс нахождения новых, действительных и потенциально полезных знаний в базах данных.

ХД - предметно-ориентированный, интегрированный, редко меняющийся, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений.

Основные характеристики хранилищ данных

- содержит исторические данные;
- хранит подробные сведения, а также частично и полностью обобщенные данные;
- данные в основном являются статическими;
- нерегламентированный, неструктурированный и эвристический способ обработки данных;
- средняя и низкая интенсивность обработки транзакций;
- непредсказуемый способ использования данных;
- предназначено для проведения анализа;
- ориентировано на предметные области;
- поддержка принятия стратегических решений;
- обслуживает относительно малое количество работников руководящего звена.

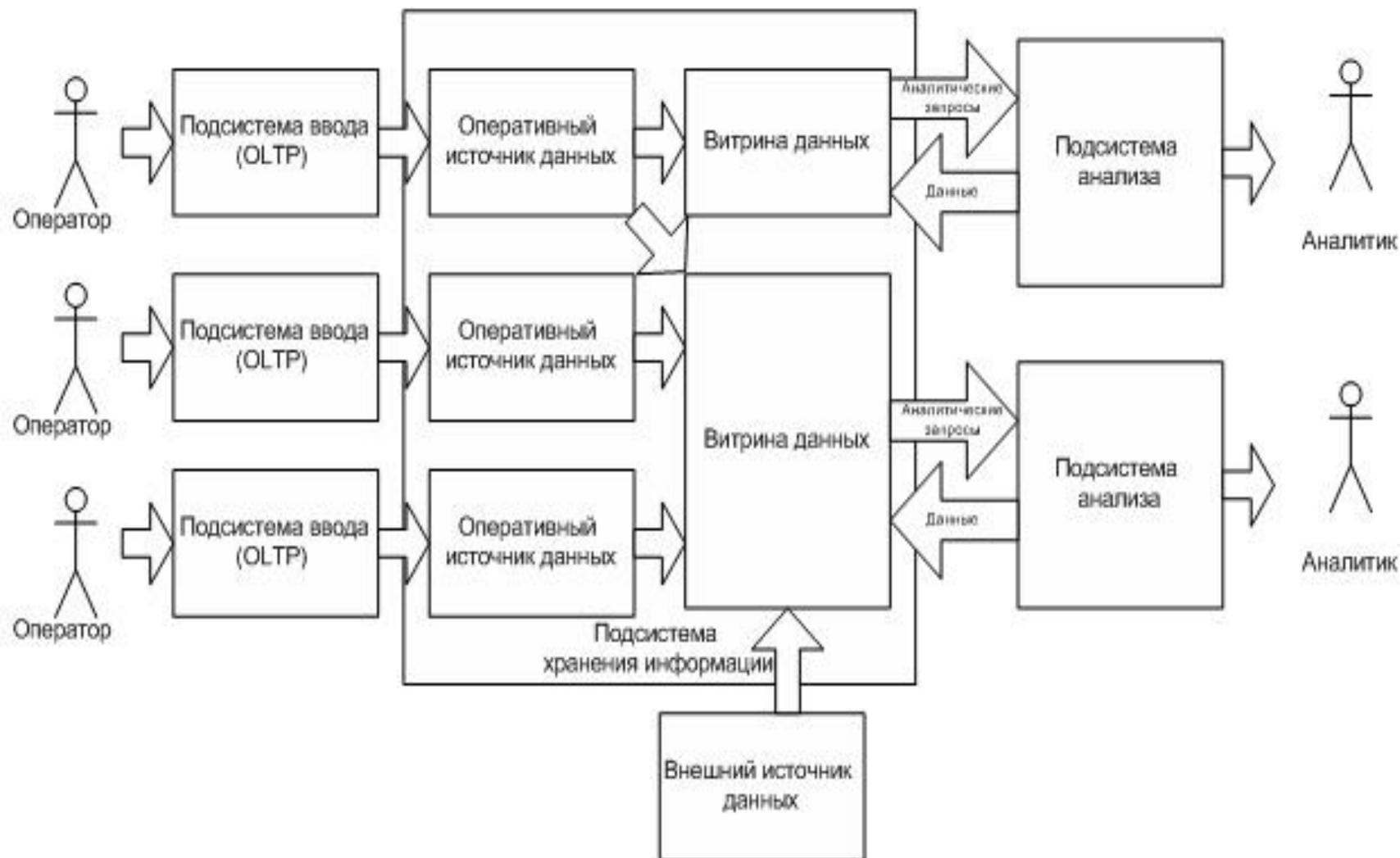
Структура СППР с физическим ХД

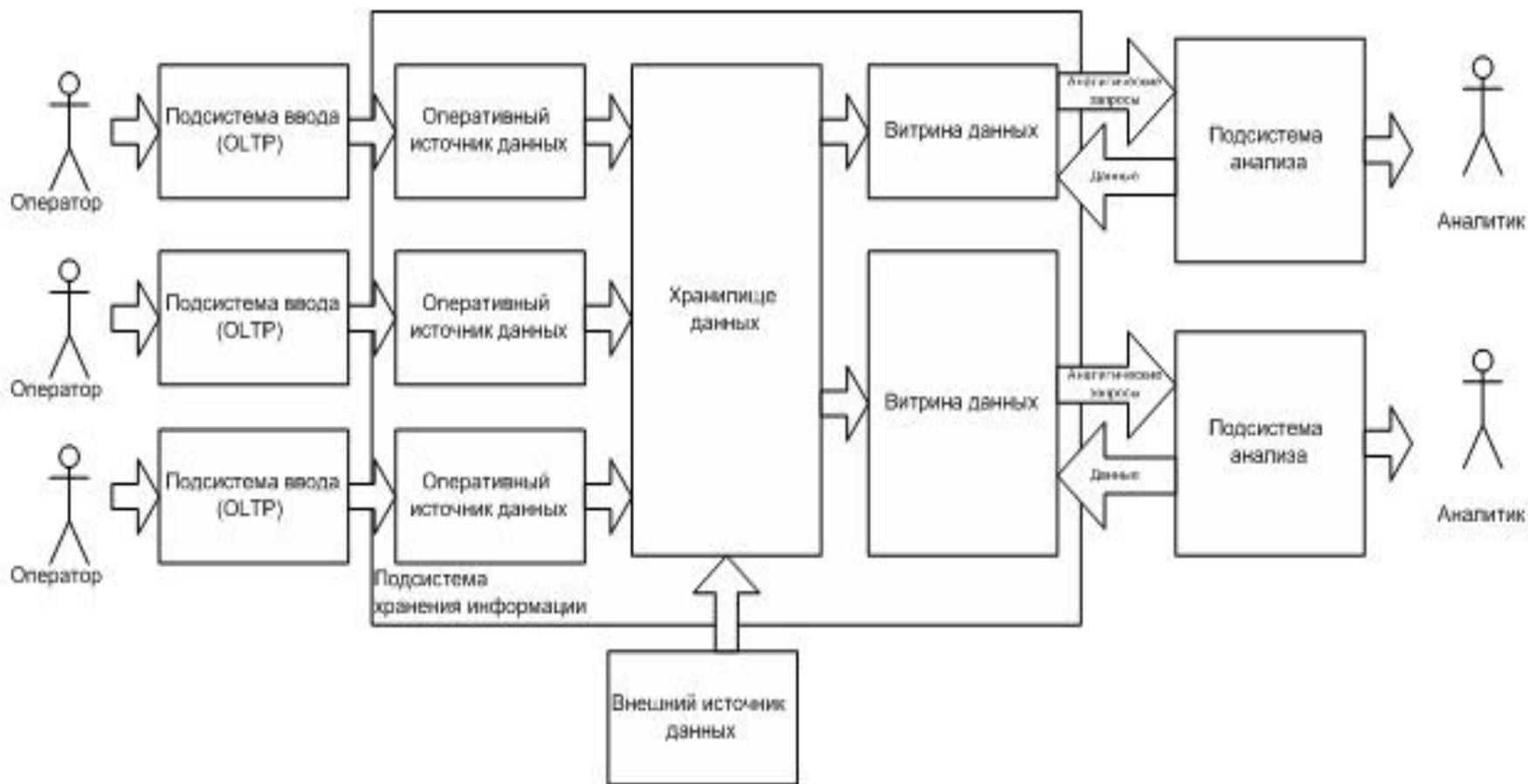


Проблемы, возникающие при создании ХД

- интеграция разнородных данных;
- эффективное хранение и обработка больших объемов данных;
- организация многоуровневых справочников метаданных;
- обеспечение информационной безопасности ХД.

Витрины данных





OLAP

OLAP (Online Analytical Processing) - технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений.

Основное назначение OLAP-систем - поддержка аналитической деятельности, произвольных запросов пользователей - аналитиков.

Цель OLAP-анализа - проверка возникающих гипотез.

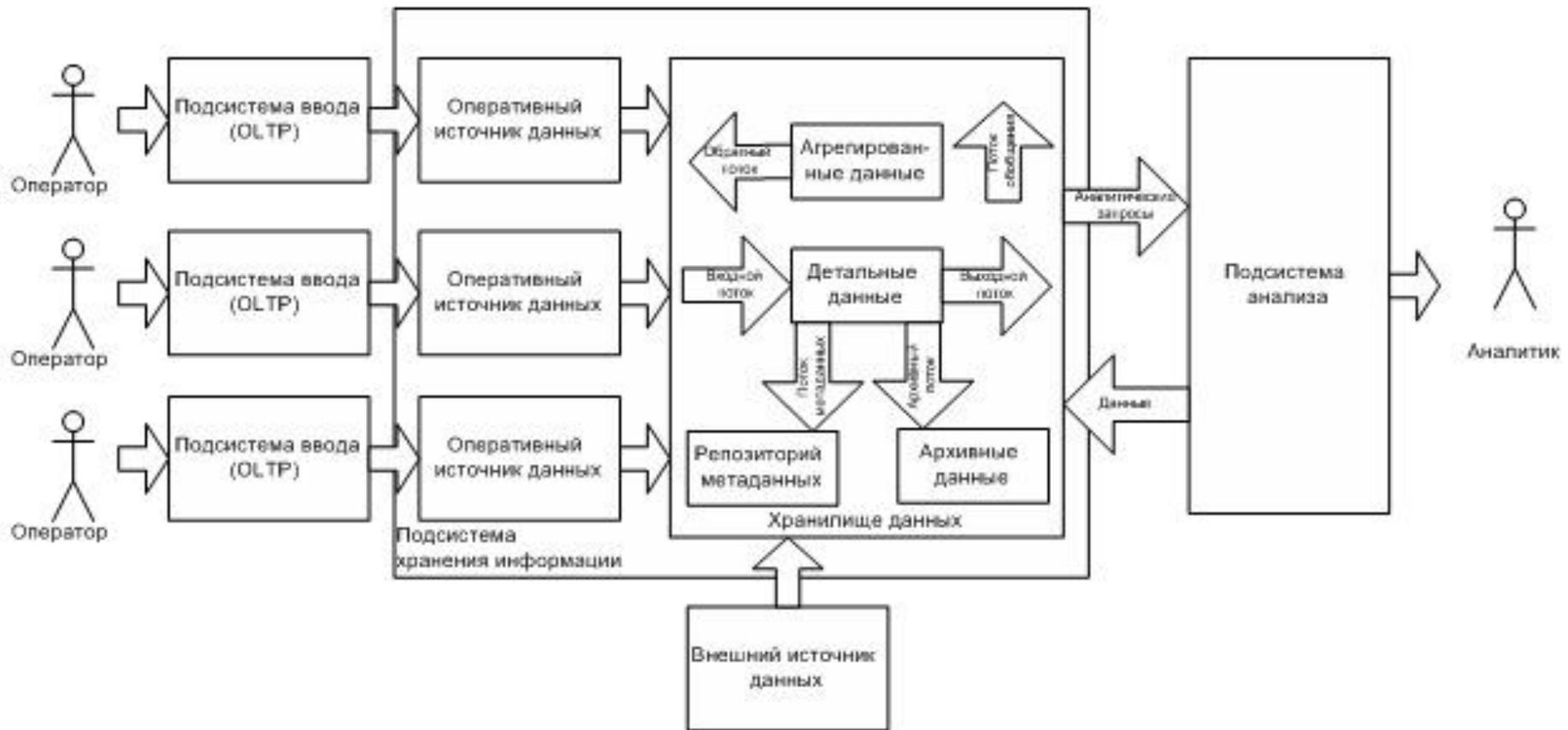
Категории данных в ХД

- детальные данные (измерения и факты)
- агрегированные (обобщенные) данные
- метаданные - данные о данных, содержащихся в ХД

Информационные потоки в ХД

- входной поток
- поток обобщения
- архивный поток
- поток метаданных
- выходной поток
- обратный поток

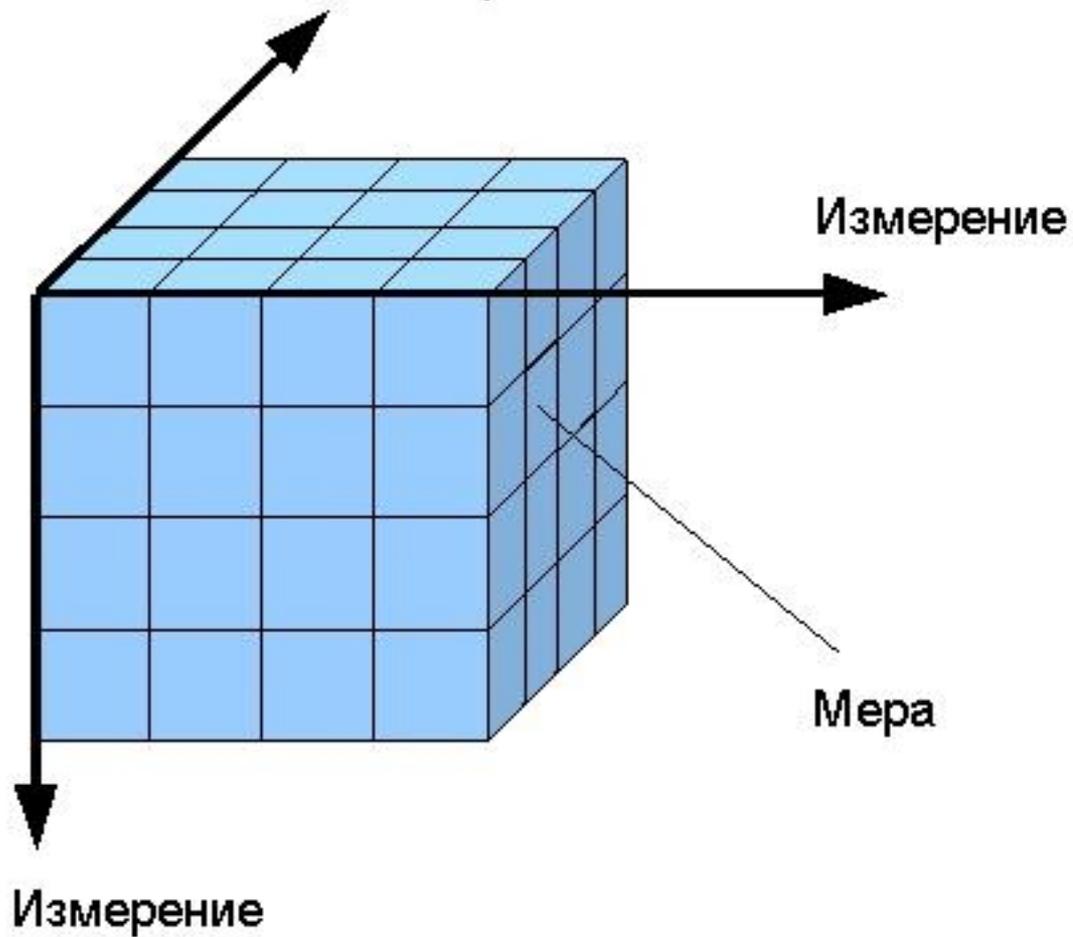
Архитектура ХД



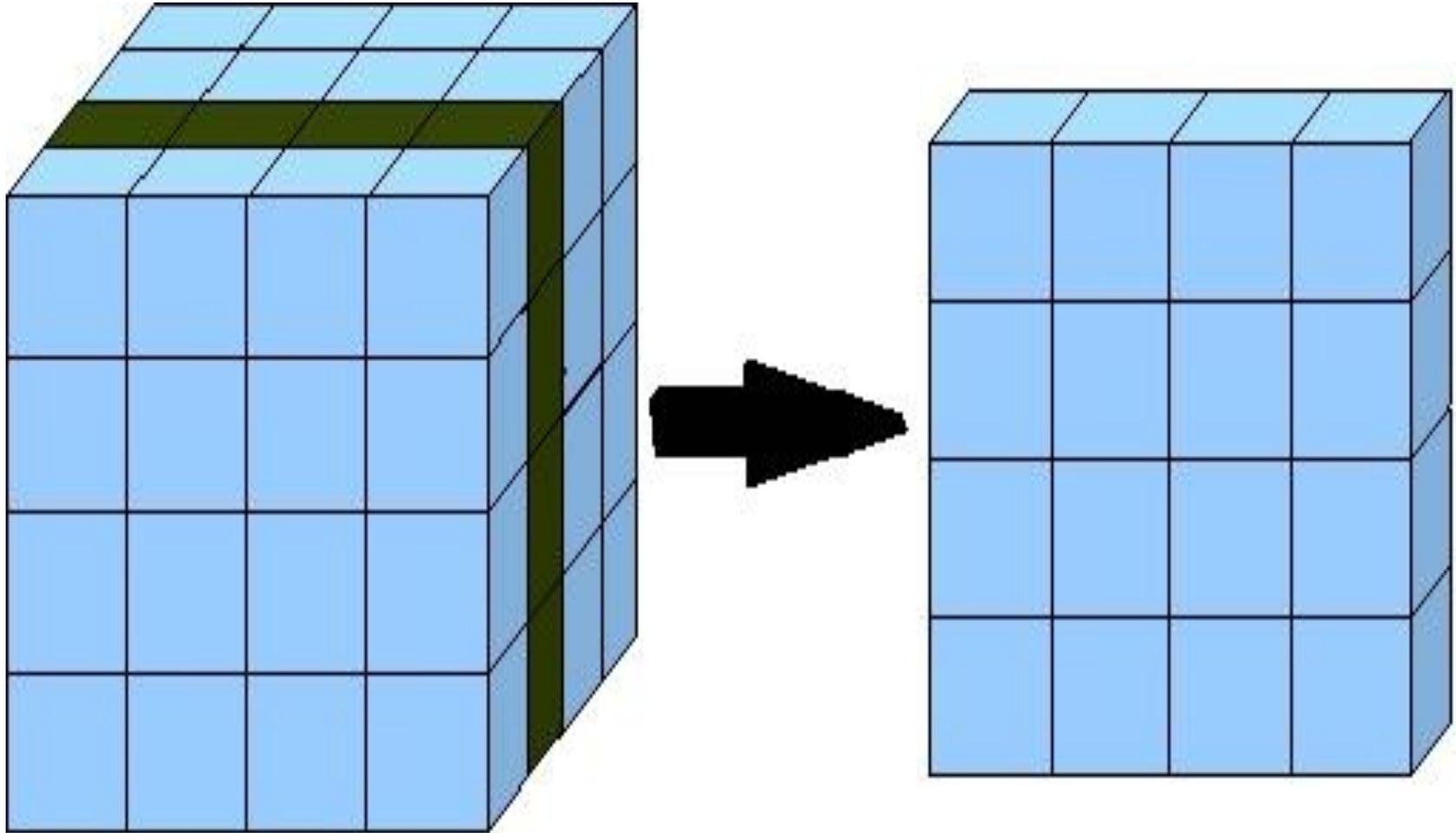
Струн

Измерение

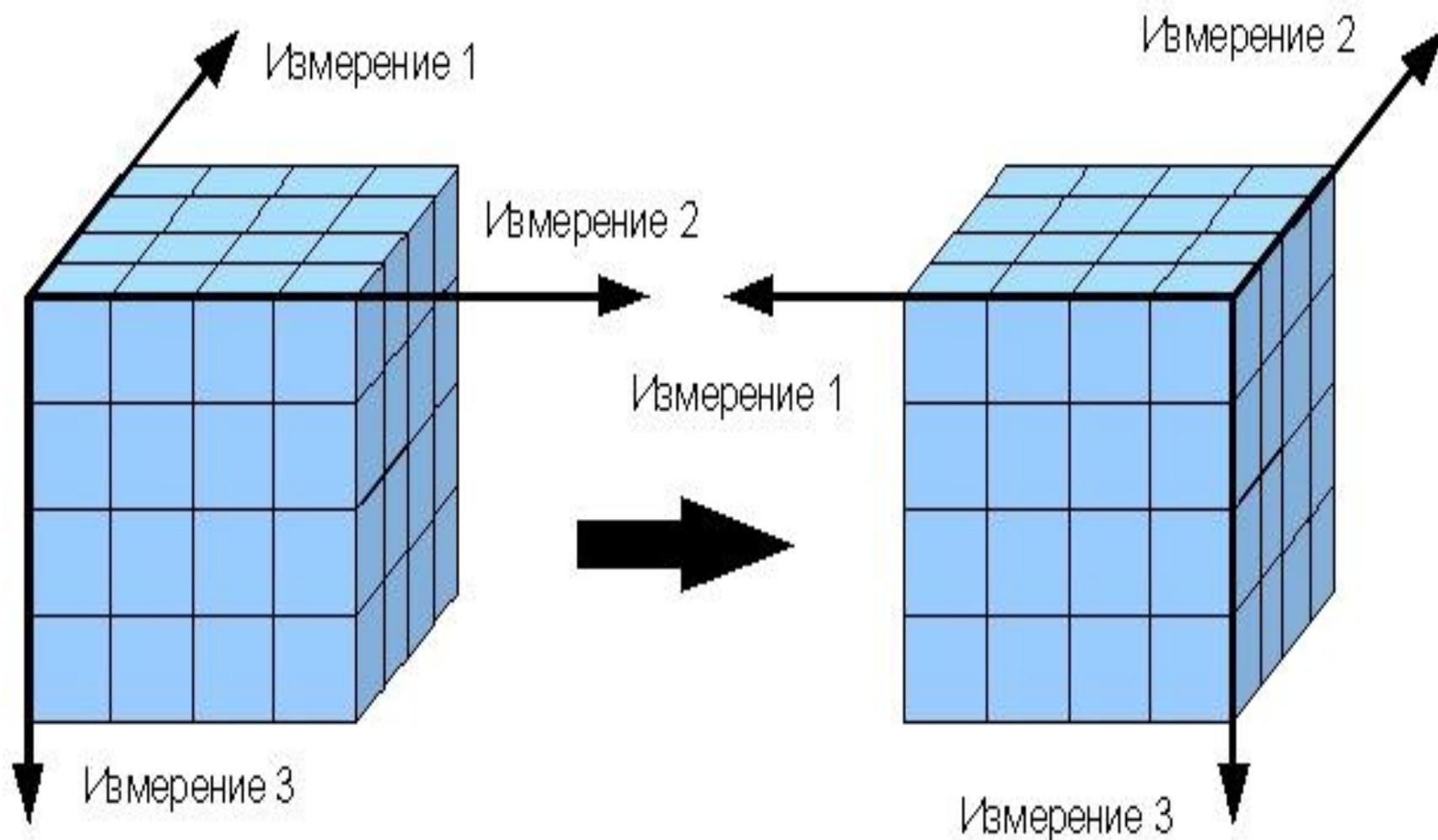
куб)



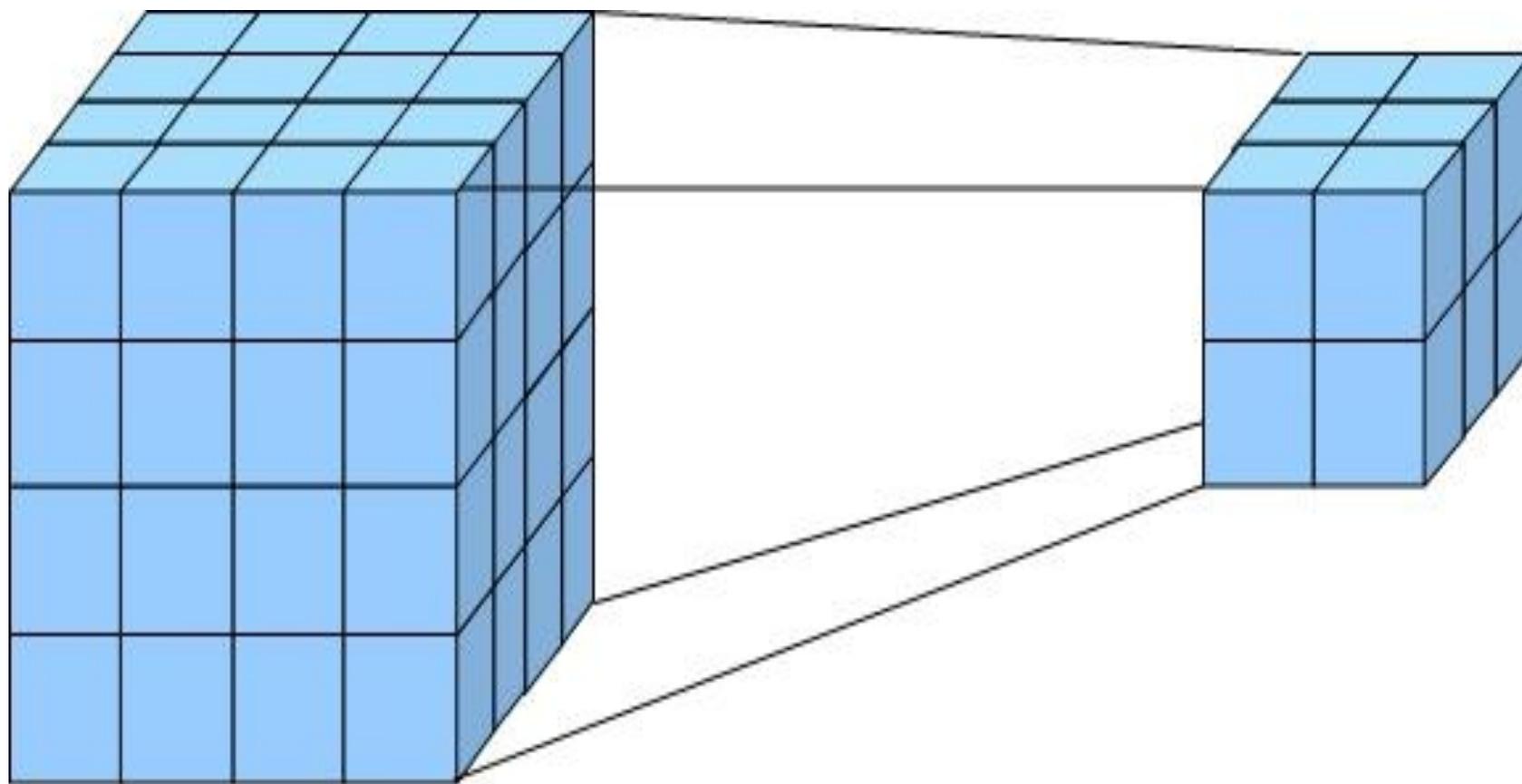
Срез



Вращение



Консолидация



Детализация

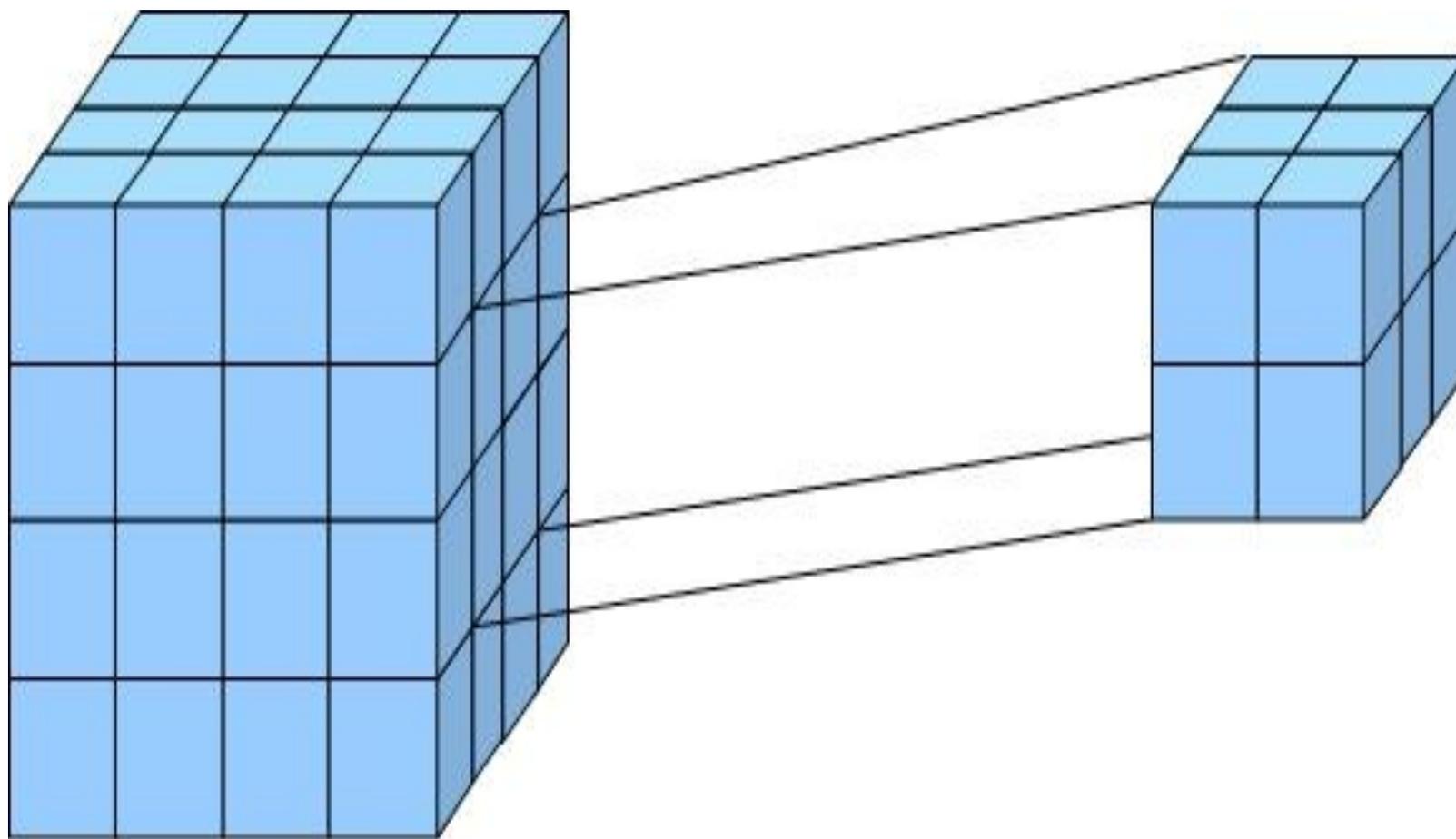
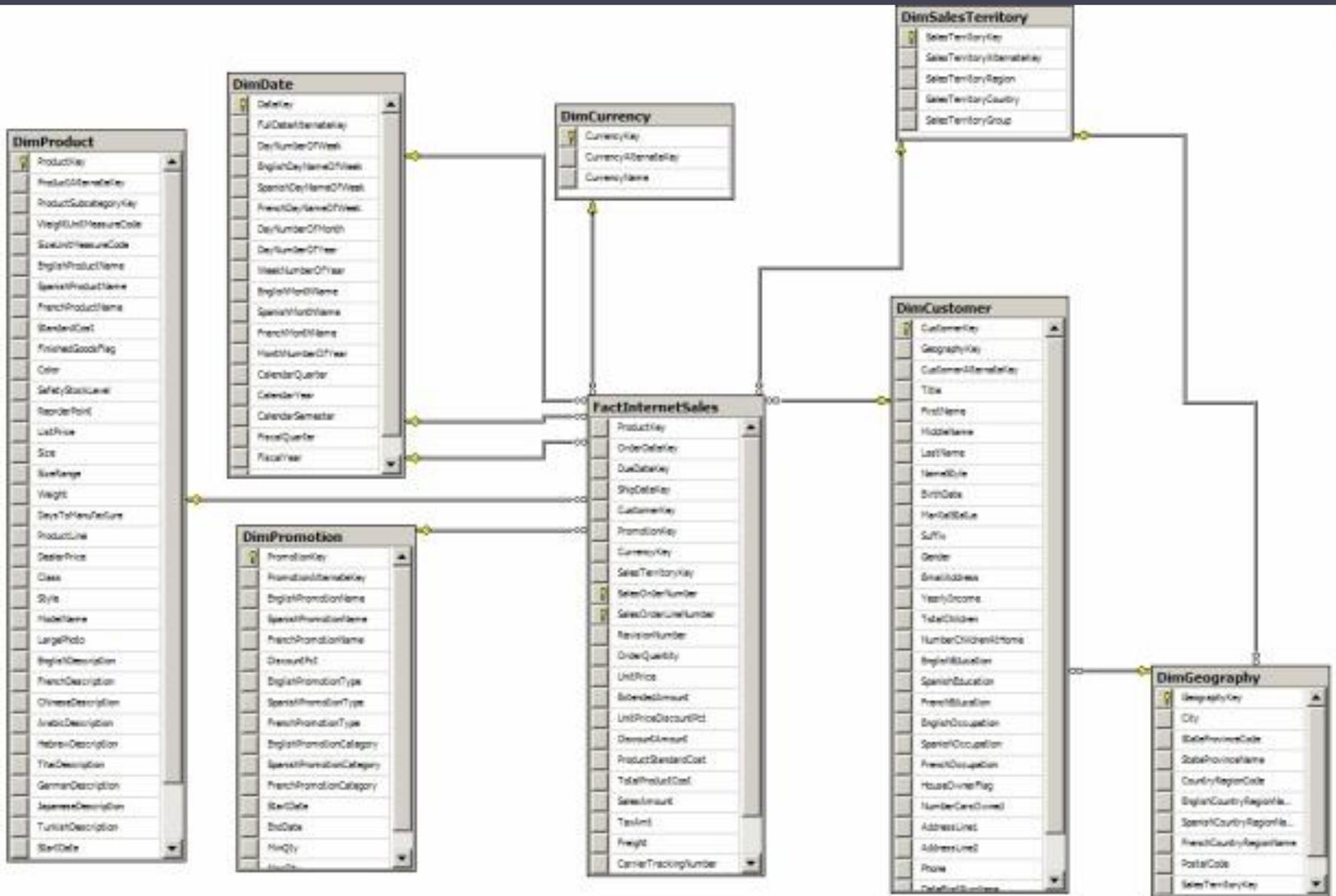


Таблица фактов

- факты, связанные с транзакциями (Transaction facts);
- факты, связанные с "моментальными снимками" (Snapshot facts);
- факты, связанные с элементами документа (Line-item facts);
- факты, связанные с событиями или состоянием объекта (Event or state facts).



Фрагмент схемы данных хранилища данных AdventureWorks

Таблицы измерений

- Таблицы измерений содержат неизменяемые либо редко изменяемые данные.
- Каждая таблица измерений должна находиться в отношении "один ко многим" с таблицей фактов.

Таблицы Фактов

	Имя столбца	Тип данных
▶	ProductKey	int
	OrderDateKey	int
	DueDateKey	int
	ShipDateKey	int
	CustomerKey	int
	PromotionKey	int
	CurrencyKey	int
	SalesTerritoryKey	int
🔑	SalesOrderNumber	nvarchar(20)
🔑	SalesOrderLineNumber	tinyint
	RevisionNumber	tinyint
	OrderQuantity	smallint
	UnitPrice	money
	ExtendedAmount	money
	UnitPriceDiscountPct	float
	DiscountAmount	float
	ProductStandardCost	money
	TotalProductCost	money
	SalesAmount	money
	TaxAmt	money
	Freight	money
	CarrierTrackingNumber	nvarchar(25)
	CustomerPONumber	nvarchar(25)

Измерений

🔑	CustomerKey	int
	GeographyKey	int
	CustomerAlternateKey	nvarchar(15)
	Title	nvarchar(8)
	FirstName	nvarchar(50)
	MiddleName	nvarchar(50)
	LastName	nvarchar(50)
	NameStyle	bit
	BirthDate	date
	MaritalStatus	nchar(1)
	Suffix	nvarchar(10)
	Gender	nvarchar(1)
	EmailAddress	nvarchar(50)
	YearlyIncome	money
	TotalChildren	tinyint
	NumberChildrenAtHome	tinyint
	EnglishEducation	nvarchar(40)
	SpanishEducation	nvarchar(40)
	FrenchEducation	nvarchar(40)
	EnglishOccupation	nvarchar(100)
	SpanishOccupation	nvarchar(100)
	FrenchOccupation	nvarchar(100)
	HouseOwnerFlag	nchar(1)
	NumberCarsOwned	tinyint
	AddressLine1	nvarchar(120)
	AddressLine2	nvarchar(120)
	Phone	nvarchar(20)
	DateFirstPurchase	date
	CommuteDistance	nvarchar(15)

Архитектура OLAP-систем

