OLAP и OLTP системы

OLTP –
оперативная транзакционная обработка данных

OLAP –
оперативная аналитическая обработка данных

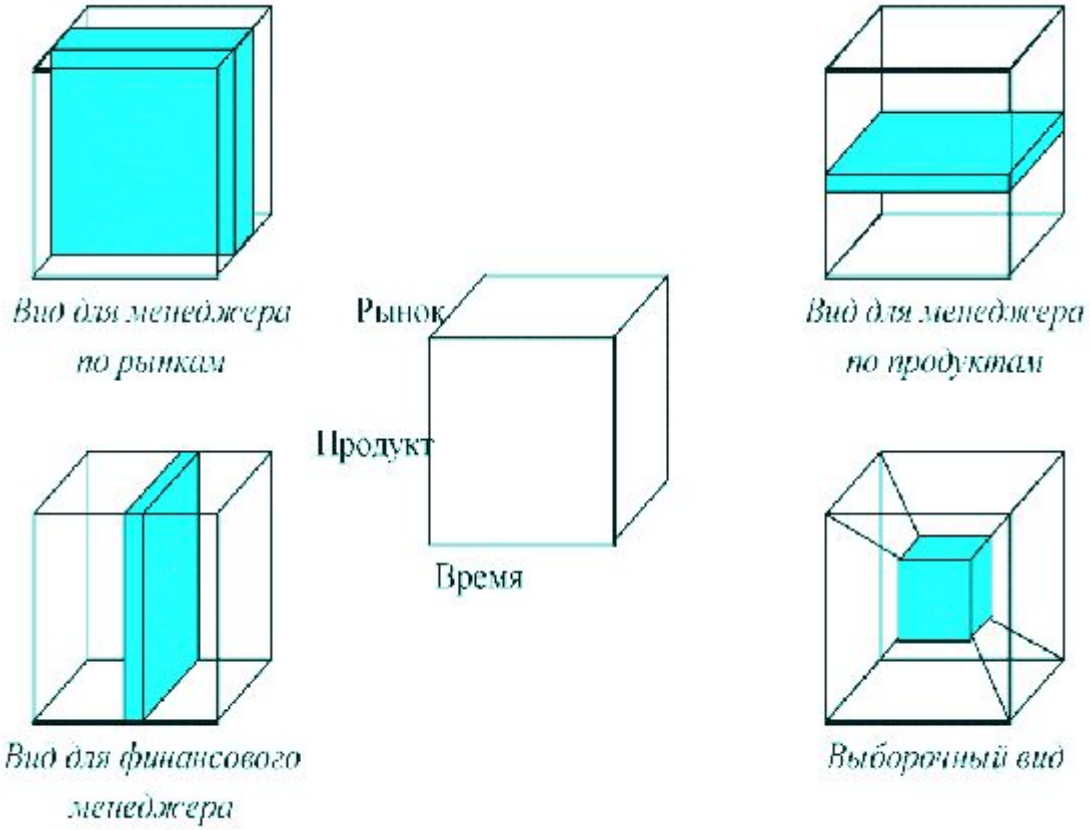
Характеристики OLTP системы

- Большой объем информации
- Часто различные БД для разных подразделений
- Нормализованная схема, отсутствие дублирования информации
- Интенсивное изменение данных
- Транзакционный режим работы
- Транзакции затрагивают небольшой объем данных
- Обработка текущих данных – мгновенный снимок
- Много клиентов
- Малое время отклика – несколько секунд

Характеристики OLAP системы

- Большой объем информации
- Синхронизированная информация из различных БД с использованием общих классификаторов
- Ненормализованная схема БД с дубликатами
- Данные меняются редко, Изменение происходит через пакетную загрузку
- Выполняются сложные нерегламентированные запросы над большим объемом данных с широким применением группировок и агрегатных функций.
- Анализ временных зависимостей
- Небольшое количество работающих пользователей – аналитики и менеджеры
- Большее время отклика (но все равно приемлемое) – несколько минут

Основной способ логического представления данных – МНОГОМЕРНЫЕ КУБЫ (OLAP – кубы)



OLAP – куб и срезы данных

	США	Канада	Мексика
Январь	20 000	4000	2000
Февраль	30 000	6000	3000
Март	50 000	10 000	5000

	США	Канада	Мексика
Unit Sales	2000	400	200
Store Sales	30 000	6000	3000
Store Cost	10 000	2000	1000

	Январь			Февраль		
	США	Канада	Мексика	США	Канада	Мексика
Unit Sales	500	100	50	500	100	50
Store Sales	7500	1500	750	7500	1500	750
Store Cost	2500	500	250	2500	500	250

Правила Кодда для реляционных БД

1. Правило информации.
2. Правило гарантированного доступа.
3. Правило поддержки недействительных значений.
4. Правило динамического каталога, основанного на реляционной модели.
5.Правило исчерпывающего подъязыка данных.
6. Правило обновления представлений.
7. Правило добавления, обновления и удаления.
8. Правило независимости физических данных.
9. Правило независимости логических данных.
10. Правило независимости условий целостности.
11. Правило независимости распространения.
12. Правило единственности.

Правила Кодда для OLAP

1. Концептуальное многомерное представление.
2. Прозрачность.
3. Доступность.
4. Постоянная производительность при разработке отчетов.
5. Клиент-серверная архитектура.
6. Общая многомерность.
7. Динамическое управление разреженными матрицами.
8. Многопользовательская поддержка.
9. Неограниченные перекрестные операции.
10. Интуитивная манипуляция данными.
11. Гибкие возможности получения отчетов.
12. Неограниченная размерность и число уровней агрегации.

Реализация OLAP

Типы OLAP - серверов

MOLAP (Multidimensional OLAP) - и детальные данные, и агрегаты хранятся в многомерной БД.

ROLAP (Relational OLAP) - детальные данные храняться в реляционной БД; агрегаты хранятся в той же БД в специально созданных служебных таблицах.

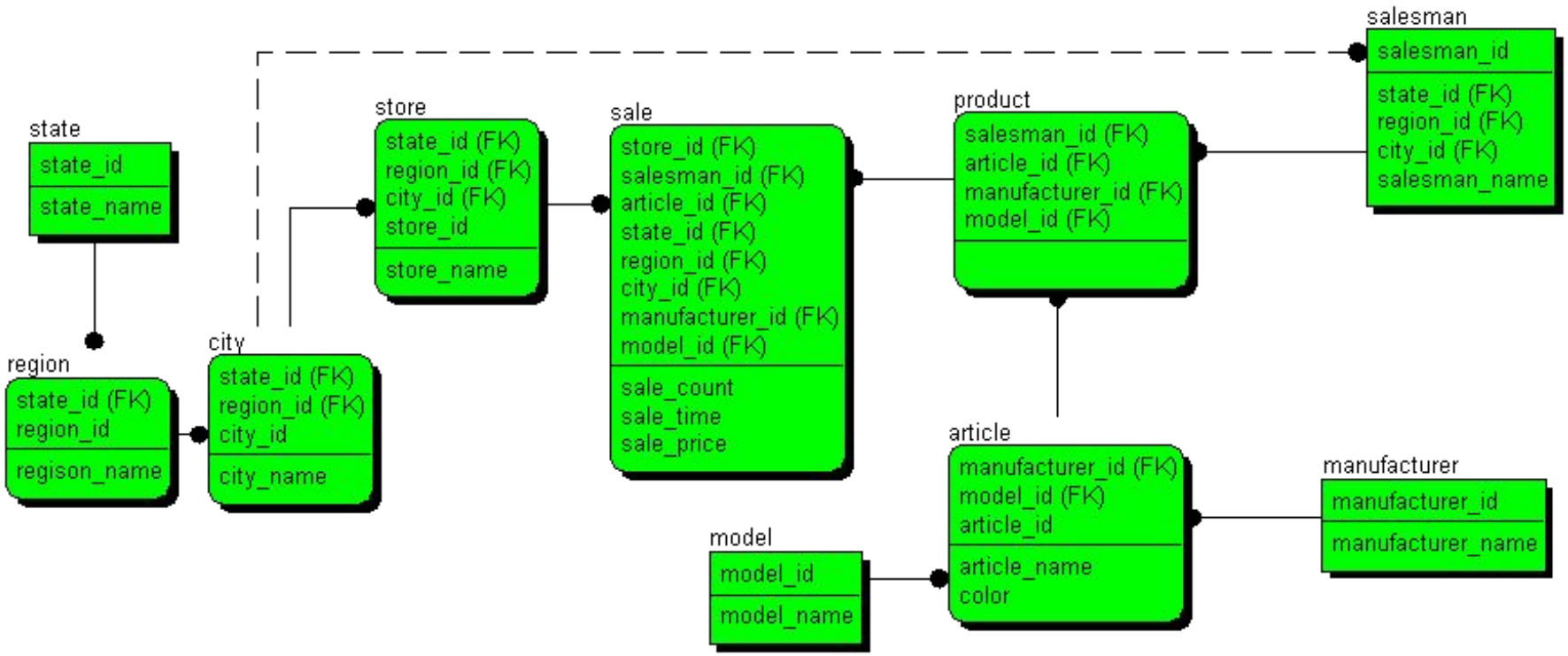
HOLAP (Hybrid OLAP) - детальные данные храняться в реляционной БД, а агрегаты хранятся в многомерной БД.

OLTP схема базы данных
Моделируются оптовые продажи на склад

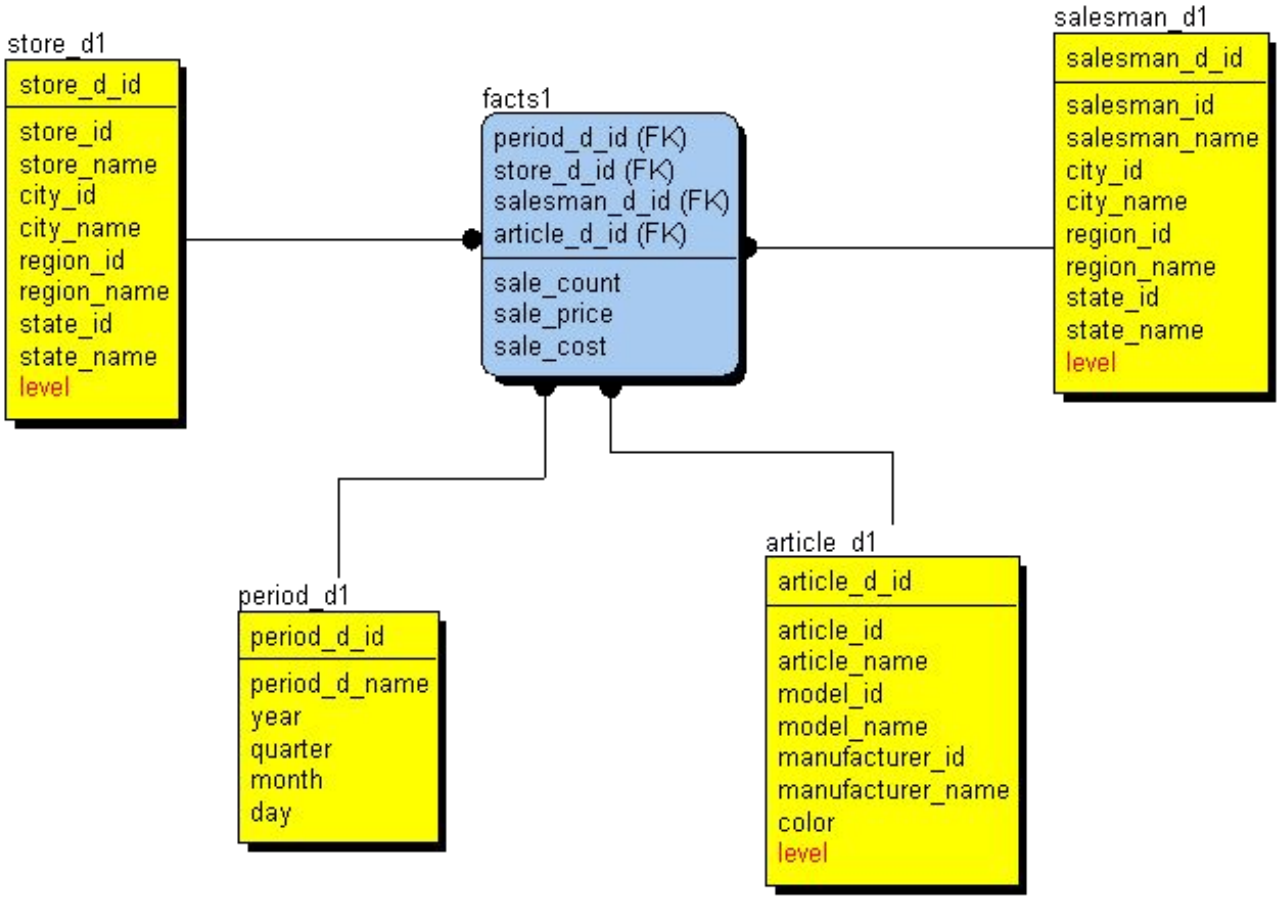
Объекты

1. Склады
2. Категории товаров (модель)
3. Производители
4. Товары
5. Продавцы
6. Оптовые продажи на склад

Оперативная схема БД оптовых продаж на склады



ROLAP – схема типа звезда



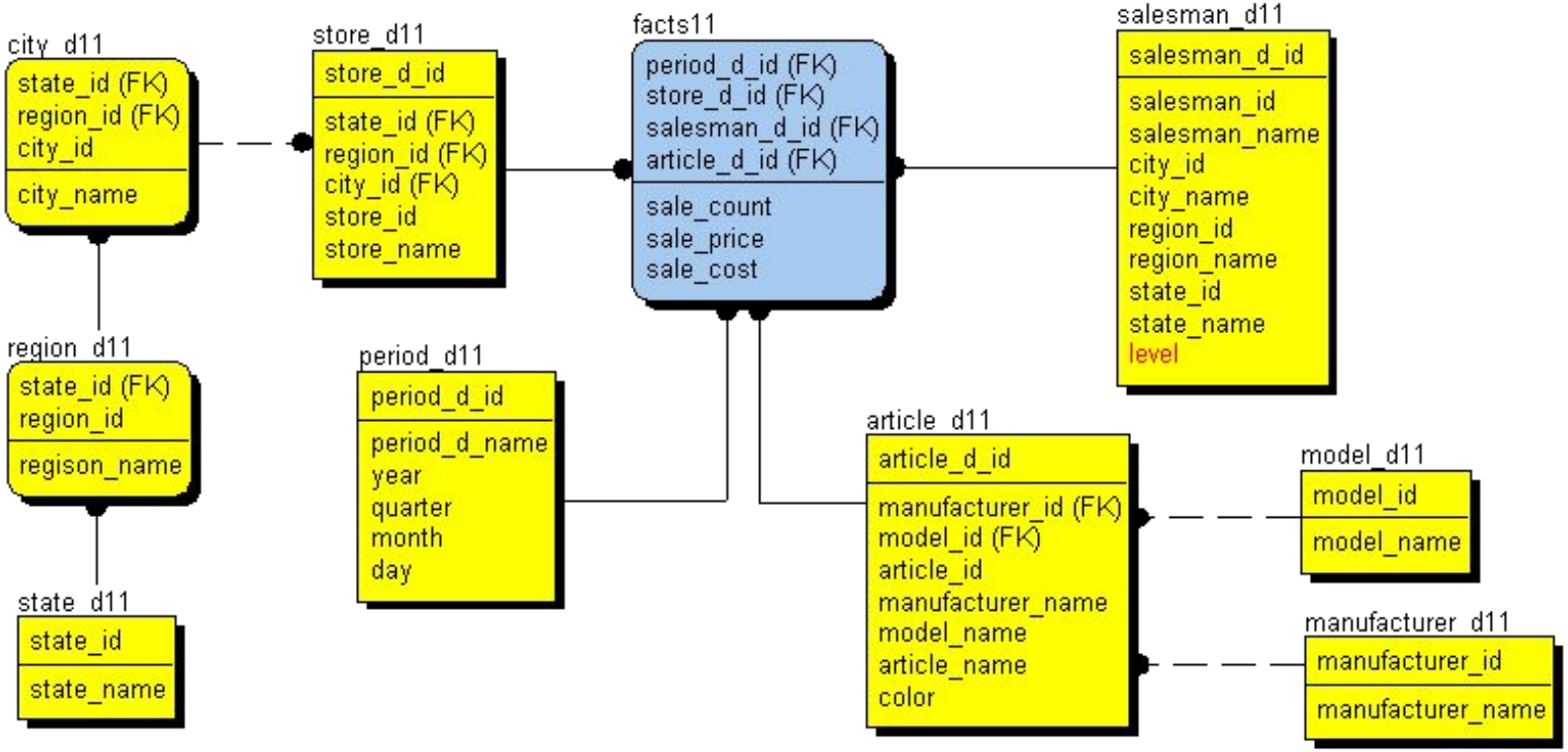
Особенности ROLAP – схемы типа звезда

1. Одна таблица фактов, которая сильно денормализована
2. Несколько таблиц измерений, которые также денормализованы
3. Первичный ключ таблицы фактов является составным и имеет по одному столбцу на каждое измерение
4. Агрегированные данные храняться совместно с исходными

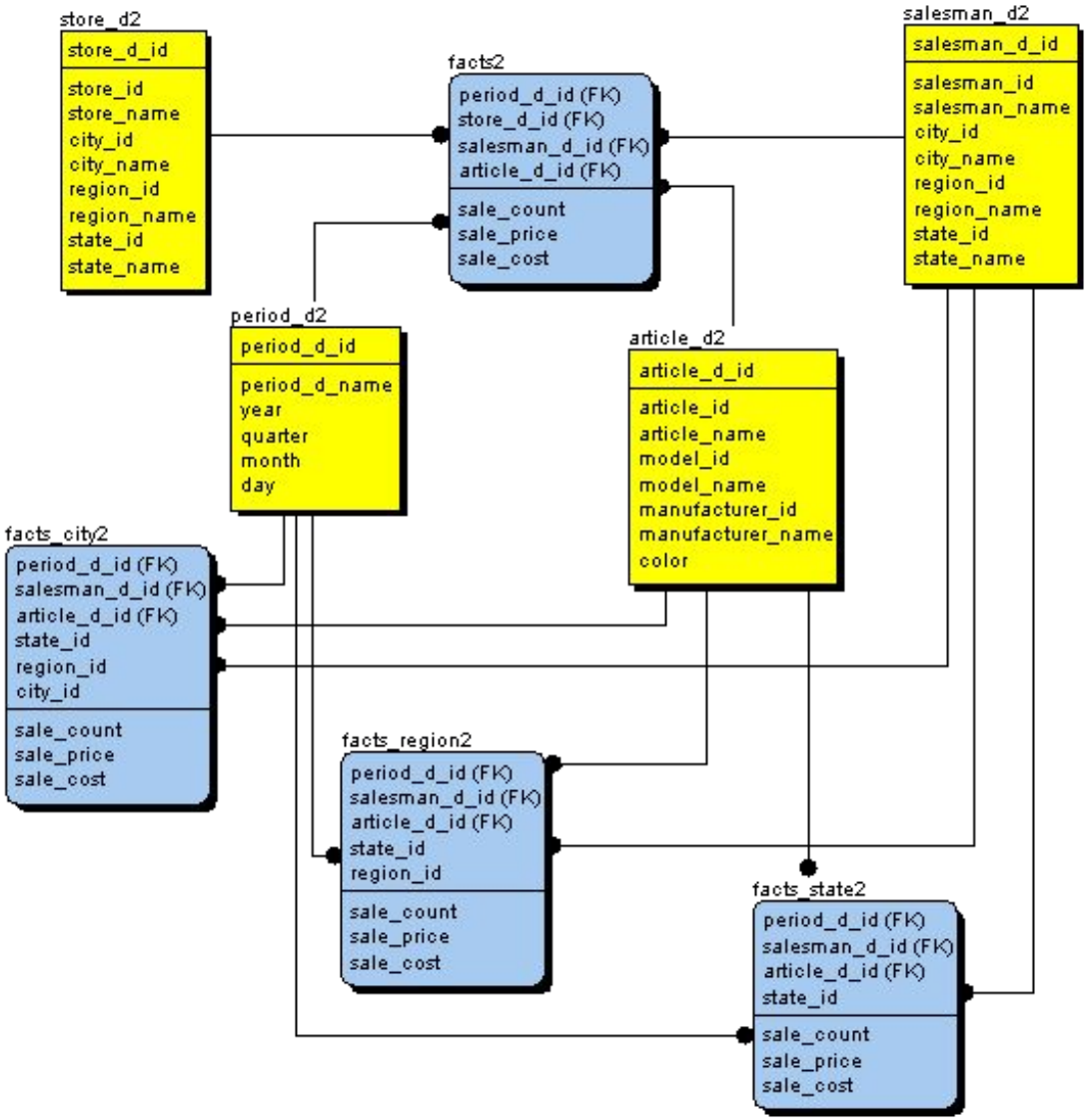
Недостатки

Если агрегаты храняться совместно с исходными данными, то в измерениях необходимо использовать дополнительный параметр – уровень иерархии

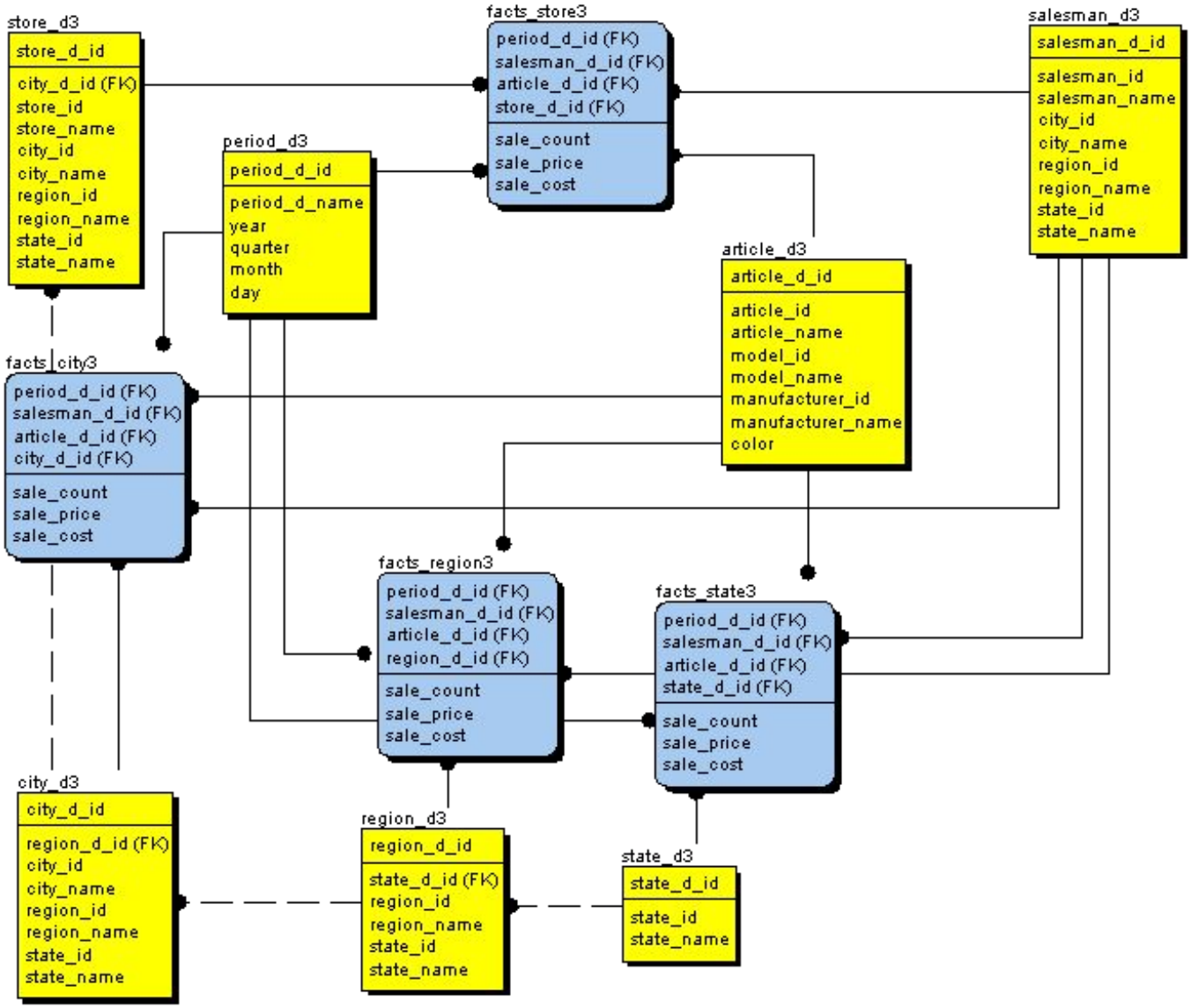
ROLAP – схема типа снежинка с нормализованными измерениями



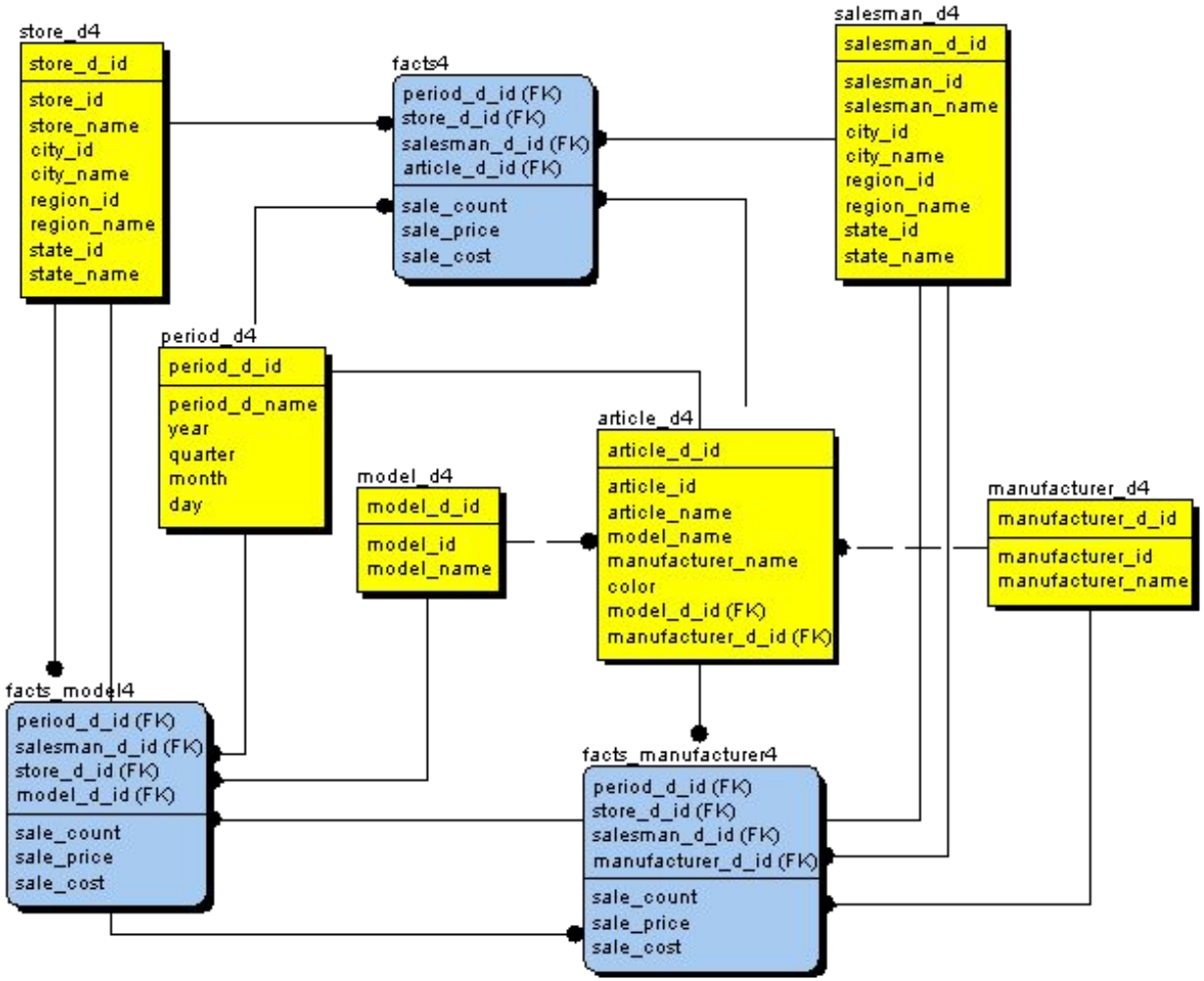
ROLAP – схема типа снежинка с выделением агрегированных таблиц



ROLAP – схема типа снежинка с выделением агрегированных таблиц и нормализованными измерениями

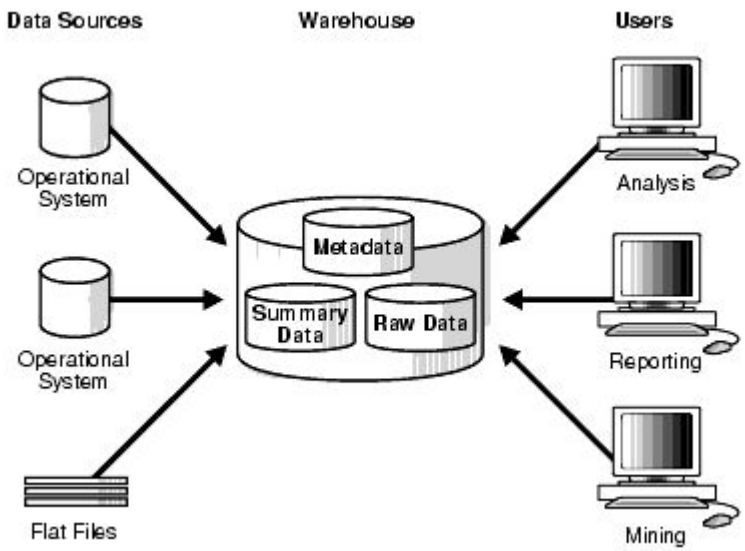


Агрегирование по производителю и модели товара



Состав хранилищ данных

1. Метаданные
2. Исходные данные
3. Предварительно просуммированные данные

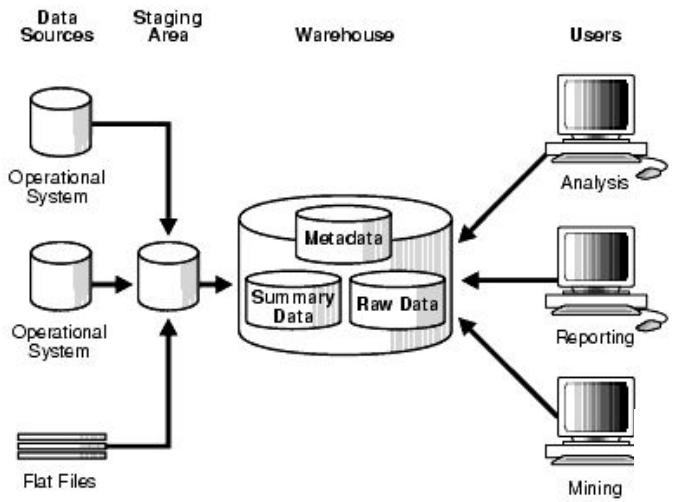


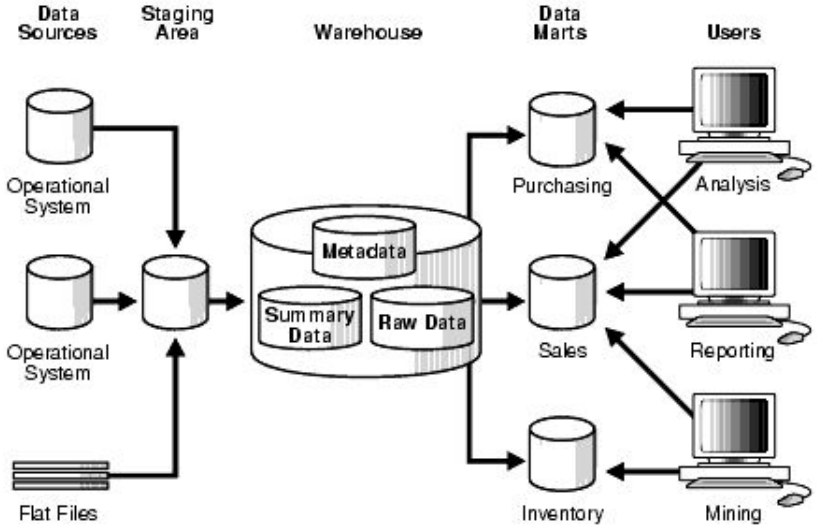
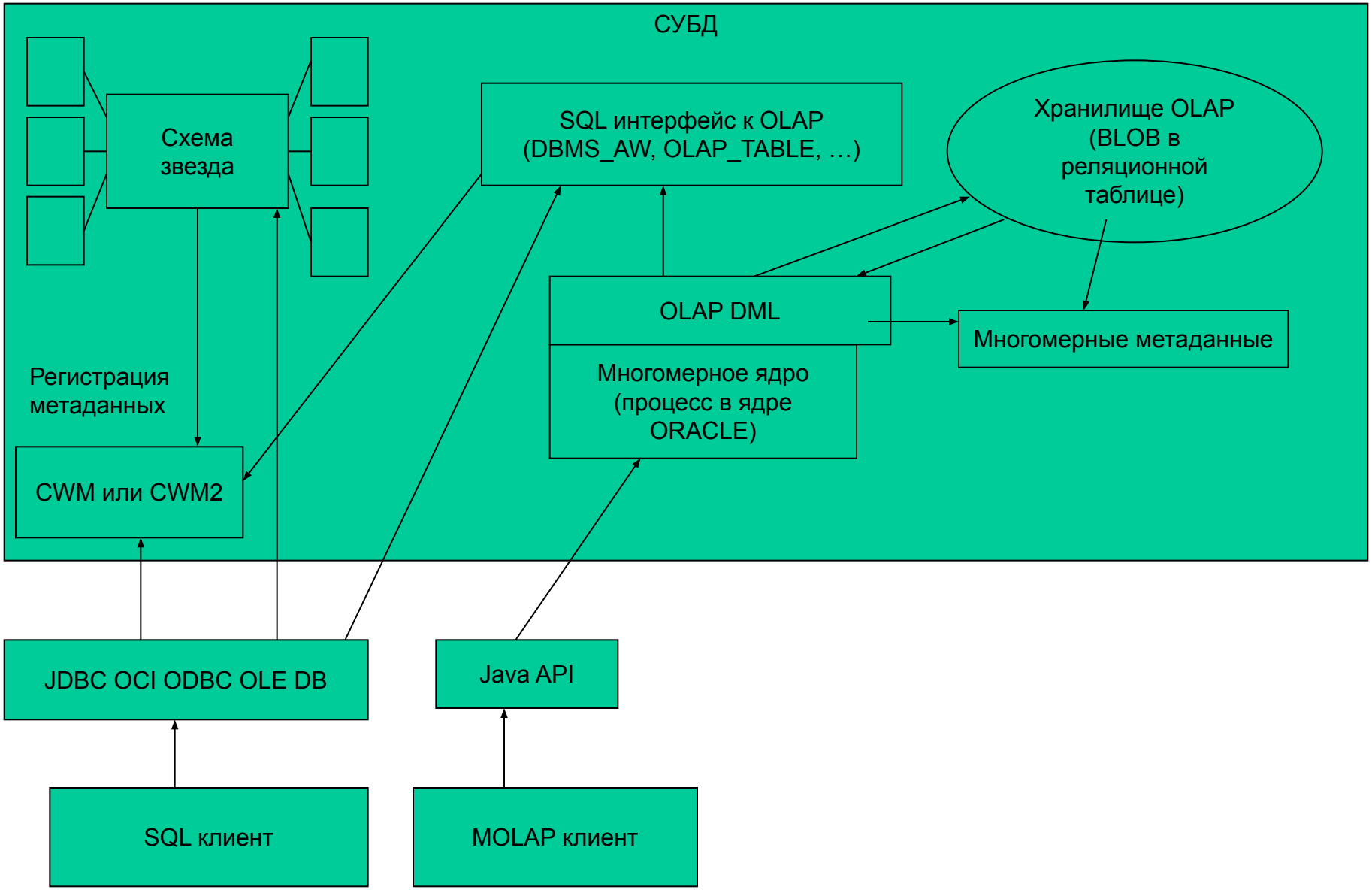
Основные метаданные OLAP

1. Куб
2. Факты
3. Измерения
    1. Уровни
    2. Иерархии
    3. Атрибуты

Общая структура хранилища данных

1. Источники данных
2. Процедуры выгрузки, преобразования и загрузки данных
3. Хранилище данных
4. Витрины данных
5. Аналитические приложения





Структура хранилища в ORACLE