**Ключевые слова:**[анализ требований](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword1), [модель данных](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword2), [БД](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword3), [OLTP](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword4), [многомерная модель](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword5), [иерархии измерений](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword8), [гранулированность](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1" \l "keyword9), [определение](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword13), [формализуемая задача](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword14), [потенциальный ключ](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1#keyword15), [outlet](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1" \l "keyword18), [quantity](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1" \l "keyword19), [granularity](https://intuit.ru/studies/courses/599/455/lecture/10168?page=1" \l "keyword21), [список](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword24), [предметной области](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword25), [анализ](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword26), [manufacturing](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword27), [TIME](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword28), [percentage](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword30), [остаток](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword36), [average](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword37), [hand](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword39), [затраты](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword40), [total](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword41), [прибыль](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword42), [процент](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword44), [прибыль от продаж](https://intuit.ru/studies/courses/599/455/lecture/10168?page=2#keyword55)

* **Тема**: Определение элементов модели ХД на основе анализа требований.
* **Задачи**:
	+ Определение измерений;
	+ Определение показателей (метрик);
	+ Определение фактов;
	+ Документирование проделанной работы.

*Анализ требований* является профессиональной задачей проектировщика ХД. При анализе требований проектировщик ХД имеет в качестве входных данных неформальные требования пользователей, модели данных для источников данных, в частности, *модель данных* *БД* *OLTP* системы, и возможно, существующую *многомерную модель* ХД, В результате анализа требований проектировщик ХД должен получить эскиз *многомерной модели* для ХД и определения элементов *многомерной модели*.

**Анализ требований проектировщиком ХД**

* В процессе анализа требований проектировщиком ХД должны быть решены следующие задачи:
	+ Определены кандидаты в показатели, факты, измерения и *иерархии измерений*.
	+ Определен уровень структуризации данных (*гранулированности*).
	+ Построен эскиз *многомерной модели*.
	+ Создан словарь предметной области для элементов модели (репозиторий метаданных).
	+ Построен эскиз *многомерной модели* ХД.

**Ключевые исходные данные**

* Корпоративная ER модель OLTP системы.
* Типовые запросы пользователей.
* Описание семантики предметной области.

**Рекомендации при решении задачи идентификации элементов модели**

* **Ориентация на запросы к ХД**. Исходя из имеющихся запросов, сначала определяются показатели, затем измерения, связанные с показателями, а затем факты.
* **Ориентация на предметную область ХД**. Исходя из элементов предметной области, сначала определяются факты, затем измерения, а потом показатели.
* **Ориентация на источники данных**. Модели данных источников данных используются как основа для выбора элементов *многомерной модели*: сначала определяются измерения, затем показатели, а потом факты.

На первом шаге проектировщик должен определить показатели, измерения с возможными иерархиями и факты. *Определение* кандидатов в элементы модели является не *формализуемой задачей*. При идентификации элементов модели могут быть использованы следующие рекомендации:

**Алгоритм действий проектировщика ХД**

* Определение кандидатов в измерения, их атрибутов и *потенциальных ключей*.
* Определение кандидатов в показатели (метрики).
* Определение кандидатов в факты из бизнес – требований.
* Определения кандидатов в факты исходя из семантического анализа предметной области.
* Документирование проделанной работы.

**Работа с корпоративной ER моделью: определение кандидатов в измерения 1**

* Определение кандидатов в измерения, их атрибутов и *потенциальных ключей*.
* Первый шаг состоит в том, чтобы вывести из модели все связанные сущности и все сущности подтипы.
	+ Это отношения N:M и подтипы.
* В ER модели нашей Компании подтипы – это сущности Product Component, Inventory, Order Line, Order, Retail Store и Corporate Sales Office.

В качестве примера мы будем использовать ER модель, которую рассматривали на прошлом семинарском занятии.

**Работа с корпоративной ER моделью: удаление подтипов**



В качестве примера мы будем использовать ER модель, которую рассматривали на прошлом семинарском занятии.

**Работа с корпоративной ER моделью: определение кандидатов в измерения 2**

* Сворачивание сущностей в конце каждого N:M отношения.
* Для каждой сущности, в том числе и выведенных подтипов, должны быть рассмотрены атрибуты, которые могли бы быть полезны в новом измерении.
* Сформировать из ключей сущностей *потенциальные ключи* измерений.
* Например, свертка Salesperson, Sales *Outlet* и Sales Region в Sales превращается в отношение Sales – Customer.

В качестве примера мы будем использовать ER модель, которую рассматривали на прошлом семинарском занятии.

**Работа с корпоративной ER моделью: кандидаты в измерения для ХД**



В качестве примера мы будем использовать ER модель, которую рассматривали на прошлом семинарском занятии.

**Бизнес -требования пользователей**

1. Какова величина среднего остатка продукции на складе и уровень запасов, при котором подается заказ, за текущий месяц для каждой модели на каждом заводе.
2. Какова величина суммарных затрат и суммарной прибыли по каждой модели, проданной сегодня, и просуммированной по отделу сбыта, типу отдела сбыта, области и по структуре продаж Компании?
3. Какова величина суммарных затрат и суммарной прибыли для каждой модели, проданной сегодня, и просуммированной по заводам и по областям?
4. Какой процент моделей получили скидки, и какие из них были проданы по факту со скидкой (в процентах) по складам (store) для всех продаж на этой неделе? В этом месяце?
5. Для каждой модели, проданной в текущем месяце, какой был процент продаж с розничной торговли, с оптовой торговли по безналичному расчету, с оптовой торговли через продавцов?
6. Какие модели и какого типа продукция не продавалась в течение последнего месяца? В течение последней недели?
7. Какие пять моделей, проданных за последний месяц, принесли наибольшую прибыль? По продажам за квартал? По суммарным затратам?
8. Какие отделы сбыта не имели продаж в течение последнего месяца для каждой модели в каждом из пяти топ-списков?
9. Какие продавцы не имели ни одной записи о продажах за последний месяц для каждой модели в каждом из трех списков 5 моделей?

Обозначим запросы через Q1, Q2, … Q9.

**Определение показателей: Запросы и показатели**

* Кандидаты в показатели можно определить исходя из бизнес - запросов. Они соответствуют элементам данных, которые пользователи используют в запросах для вычисления эффективности и поведения бизнес-процессов и объектов деятельности.
* Хорошими кандидатами в показатели являются числовые атрибуты, при определении которых используется агрегатные функции.

**Пример**

* Если потенциальные пользователи Компании часто используют запросы типа 1, 2 и 3:
	+ Определить Суммарные затраты (Total Cost)
	+ Определить средний остаток продукции на сладе (Average *Quantity* on Hand)
	+ Определить Суммарная прибыль (Total Revenue),
* то кандидатами на показатели могут быть
	+ Средний остаток продукции на сладе (Average *Quantity* on Hand)
	+ Суммарные затраты (Total Cost)
	+ Суммарная прибыль (Total Revenue),

**Определение уровня структуризации показателей**

* Уровень структуризации показателей (Measure *Granularities*) в *многомерной модели* может быть определен, как наименьший уровень детализации, используемый для представления показателя в модели.
* Например, значение атрибута Остаток продукции на складе может сохраняться в ХД за день, за месяц или квартал. Этот атрибут может также рассматриваться по продукции или категории продукции или в единицах упаковки.

**Определение уровня структуризации показателей**

* Показатель обычно связан с несколькими измерениями. Уровень структуризации показателя определяется комбинацией детализации каждого из этих измерений. Различные показатели могут иметь одинаковый уровень структуризации.
* Например, показатели Суммарные затраты (Total Cost) и Суммарная прибыль (Total Revenue) связаны с операцией продажи, и поэтому они имеют одинаковый уровень структуризации. Такие показатели являются претендентами на другой элемент *многомерной модели* - факты.
* Определение правильного уровня структуризации показателей в модели ХД является одной из главных задач проектировщика ХД. Уровень структуризации показателей определяет уровни, на которых пользователь может решать свои аналитические задачи.
* Например, 5 уровней структуризации данных обеспечивает выполнение анализа информации на 5 уровнях ее детализации. Ясно, что при увеличении уровня структурированности данных объем сохраняемой в ХД информации увеличивается, что влияет как на размер ХД, так и на производительность обработки запросов.