

Лабораторна робота 3

Тема: Мапування потоків даних для ETL процесу між БД транзакційного типу та OLAP сховищами даних

Теоретичні відомості

Багатовимірне моделювання є методом моделювання та візуалізації таких даних як множини числових або лінгвістичних показників або параметрів (measures), які описують загальні аспекти діяльності організації.

Метод багатовимірного моделювання базується на таких основних поняттях: факти, атрибути, виміри, параметри (метрики), ієрархія, гранулювання.

Факт (fact) - це набір пов'язаних елементів даних, що містять метрики і описові дані. Кожен факт звичайно являє елемент даних, чисельно описує діяльність організації, бізнес-операцію чи подію, що може бути використаний для аналізу діяльності організації або бізнес-процесів. У ХД факти зберігаються в базових таблицях реляційної БД. Наприклад, вартість товару, кількість одиниць товару і т.д.

Атрибут (Attribute) - це опис характеристики реального об'єкта предметної області. Як правило, атрибут містить задалегідь відоме значення, що характеризує факт. Зазвичай атрибути представляються текстовими полями з дискретними значеннями. Наприклад, габарити упаковки товару, запах товару.

Вимір (dimension) - це інтерпретація факту з деякою точки зору в реальному світі. Виміри, подібно атрибутам, містять текстові значення, які сильно пов'язані за змістом між собою. Зазвичай виміри це осі багатовимірного простору, точками якого є пов'язані з ними факти. У багатовимірній моделі кожен факт пов'язаний з однією або декількома осями. Виміри зазвичай представляють нечислові, лінгвістичні змінні, такі як філії організації, співробітники організації, покупці і т.д.

Параметр, метрика або показник (measure) - це числова характеристика факту, який визначає ефективність діяльності або бізнес-дії організації з погляду вимірювання. Як правило, метрика містить задалегідь не відоме значення характеристики факту. Конкретні значення метрики описуються за допомогою змінних. Наприклад, нехай метрикою є чисельне вираження продажів товару в грошах, кількість проданих одиниць товару тощо. Метрика визначається за допомогою комбінації елементів виміру, і, таким чином, являє факт.

Гранулювання (Granularity) - це рівень деталізації даних, що зберігаються в ХД. Наприклад, щоденні обсяги продажів.

З погляду взаємозв'язку вимірювань і фактів останні можна розбити на наступні класи:

- адитивні факти (Additive facts). Факт називається адитивним, якщо його має сенс використовувати з будь-якими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, дискретні числові показники активності діяльності, такі як кількість продажів, обсяг продажів тощо;
- полуадитивні факти (Semiadditive facts). Факт називається полуадитивним, якщо його має сенс використовувати спільно з деякими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, числові показники інтенсивності, такі як залишок на рахунку, рівень запасів на складі тощо;

- неаддитивну факти (Non-additive facts). Факт називається неаддитивну, якщо його не має сенсу використовувати спільно з яким-небудь виміром для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, вимірювання кімнатної температури;

- числові заходи інтенсивності (Numerical Measures of Intensity). Факт називається числовий мірою інтенсивності, якщо він, будучи неаддитивним за часом, допускає агрегацію і підсумовування по деякому числу часових періодів. Наприклад, залишок на рахунку.

Таблиці фактів поділяють на три основні категорії, залежно від рівня деталізації фактів (гранулированности).

- Транзакційна таблиця фактів. У такій таблиці фактів зберігають факти, які фіксують певні події (транзакції). Це факти, що описують кожну подію бізнесу. Наприклад, продажі товару.

- Таблиця фактів періодичних моментальних знімків У такій таблиці збирають факти, що фіксують поточний стан певного напрямку бізнесу. Це факти, які описують поточний стан певного напрямку бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі організації на певну дату (щоденно).

- Таблиця фактів кумулятивних моментальних знімків У такій таблиці збирають факти, що фіксують деякий підсумковий стан певного напрямку бізнесу на поточний момент часу. Це факти, які описують проміжні підсумки діяльності організації за певним напрямком бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі цього року на певну дату.

Основними характеристиками таблиці вимірів є наступні.

1. Таблиці вимірів містять дані про деталізації фактів.

2. Таблиці вимірів містять описову інформацію про числових значеннях в таблиці фактів, тобто вони містять атрибути фактів.

3. Як правило, денормалізовані таблиці вимірів містять велику кількість полів.

4. Таблиці вимірів містять зазвичай значно менше рядків, ніж таблиці фактів.

5. Атрибути таблиць вимірів зазвичай використовуються при візуалізації даних у звітах і запитах.

Схема "зірка" має одну таблицю фактів і кілька таблиць вимірів. Таблиці вимірів є денормалізованими.

Схема "сніжинка" має одну таблицю фактів і кілька нормалізованих таблиць вимірів.

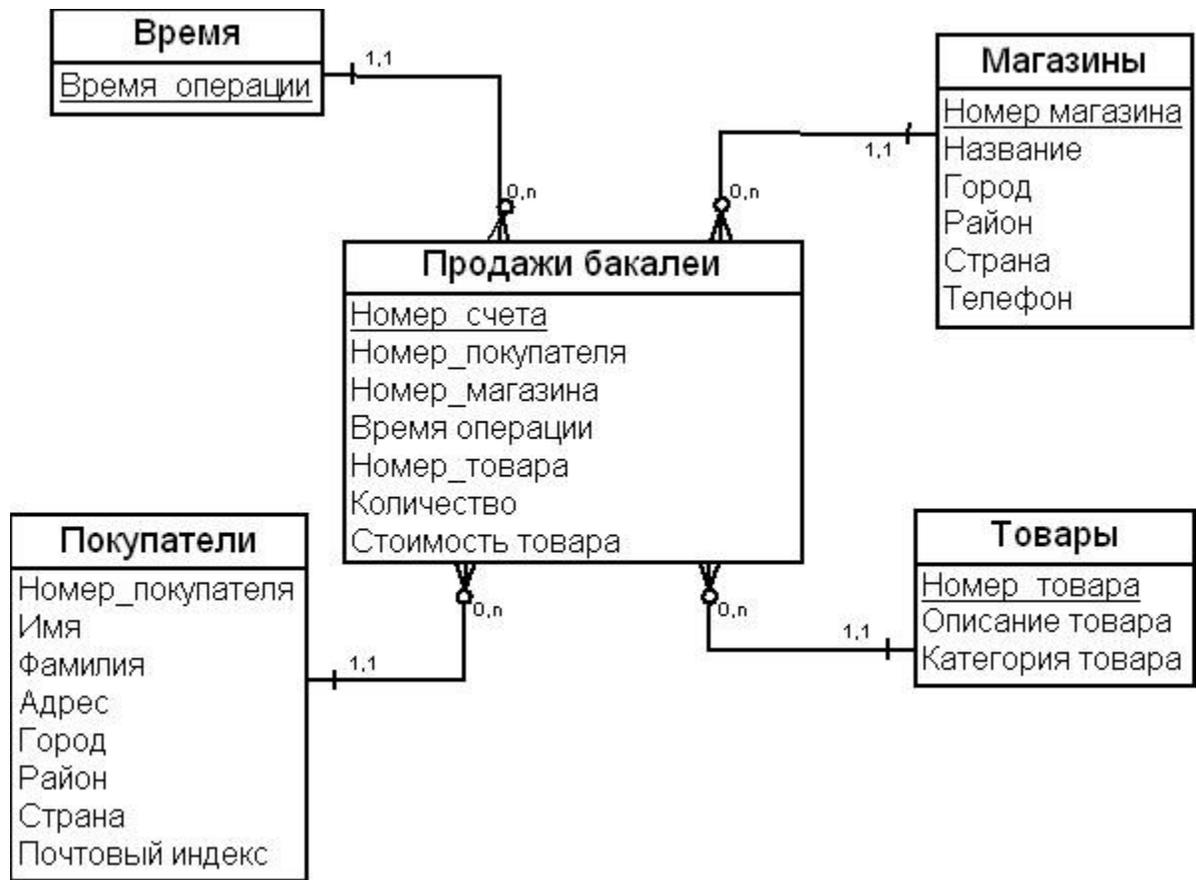


Рис.4. Схема зірка

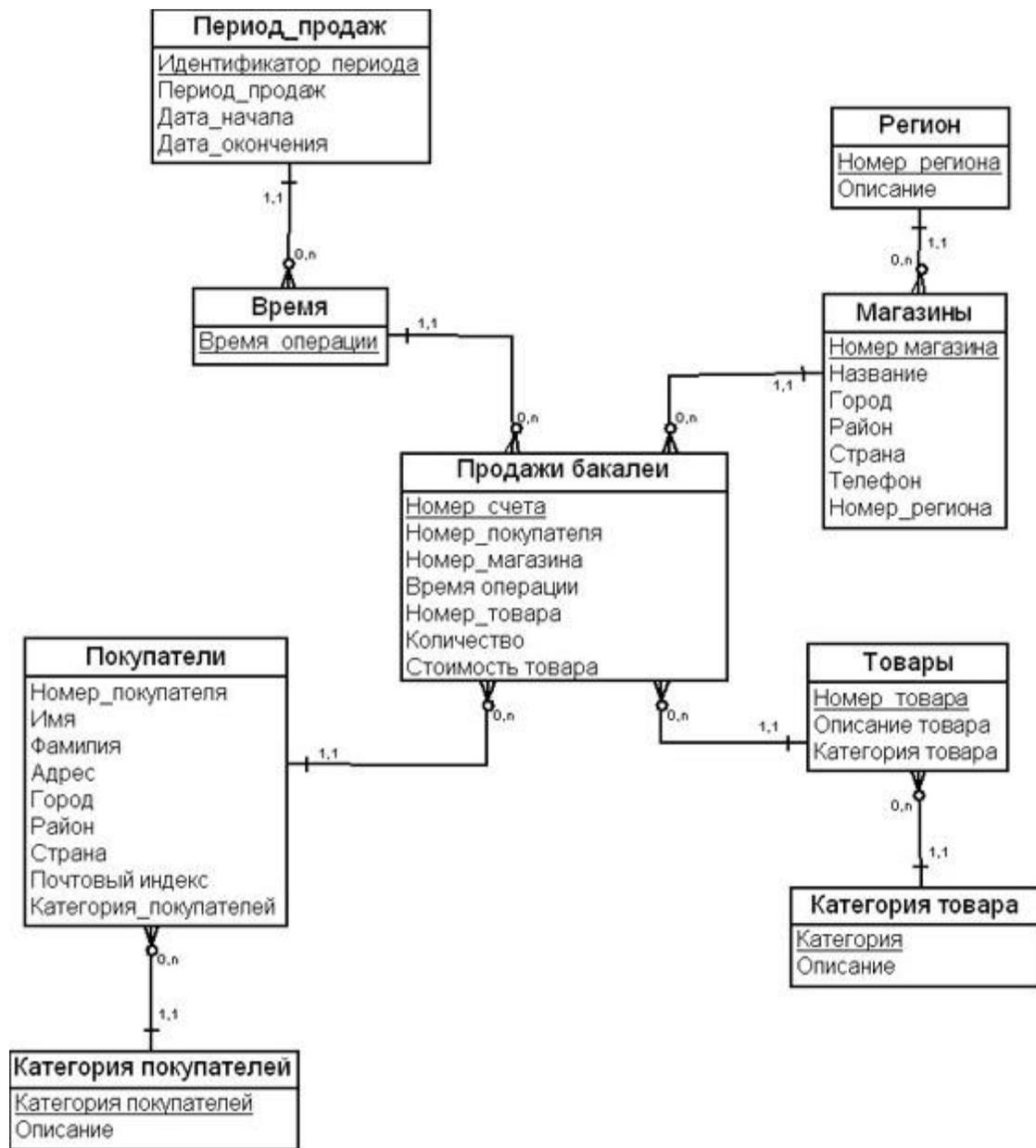


Рис.5. Схема сніжинка

Завдання на лабораторну роботу

1. За допомогою CASE засобів для проектування за стандартом IDEF1X (SQLDBM, ERWin) визначити структуру БД.
2. Спроекувати на основі топології «Зірка» сховище даних (Data WereHouse) за індивідуальним варіантом.
3. Побудувати DFD-діаграму для ETL-процесу завантаження даних.
4. Визначити потребу у проміжних сховищах для трансформації даних та в процедурах очистки даниї.

Рекомендації до виконання роботи

1. Сбор требований к хранилищу данных.

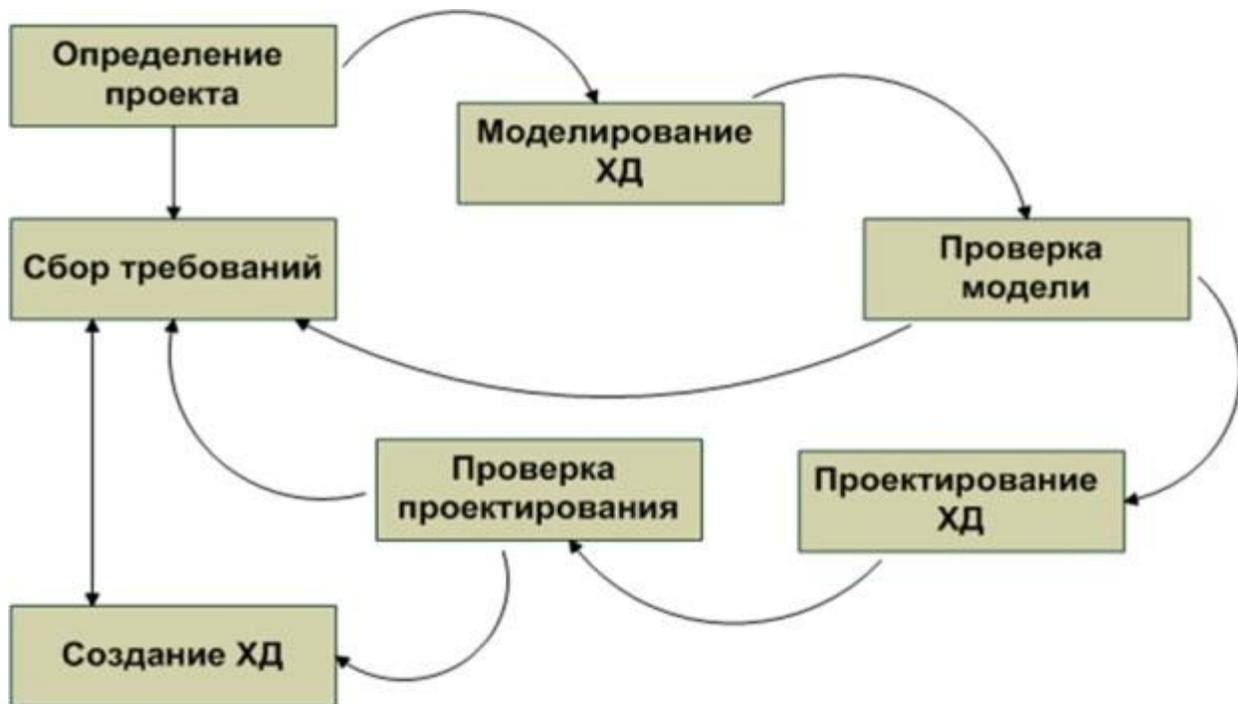
- **Задачи:**

- Анализ деятельности компании;
- Описание бизнес-процессов;
- Формулировка требований к хранилищу данных;
- Сбор данных из источников

Рассмотрим основные вопросы управления проектом разработки и создания ХД. В первую *очередь* мы затронем те аспекты жизненного *цикла* разработки ХД, которые оказывают наибольшее влияние на этап проектирования.

Мы будем строить ХД на примере решения задачи анализа продаж некоторой гипотетической Компании.

Простая модель процесса разработки хранилища данных



В основу методологии проектирования хранилищ данных может быть положена простая *математическая модель* процесса разработки программного изделия.

Модель процесса представляет собой обычную модель бизнес-процессов проектирования хранилищ данных, которая содержит укрупненные бизнес-задачи разработки ХД.

Диаграмма жизненного цикла процесса является отправной точкой такой модели.

Что должен обязательно делать проектировщик хранилища данных.

Уметь собрать требования к ХД совместно с бизнес - аналитиком.

Уметь анализировать требования, возможно с бизнес - аналитиком.

Уметь проверять требования.

Создать эскиз *многомерной модели* на основе собранных требований.

Определение проекта создания ХД: Оценка ситуации в целом

- ХД заказывает организация - Компания, которая занимается производством и продажей телефонов.
- Компания создана недавно, около 10 лет назад. Сначала около 7 лет она занималась только производством телефонов, сбыт был организован через дистрибьютерскую сеть.
- Спрос на телефоны вырос, и Компания создала свою собственную сеть магазинов для продажи телефонов.
- Вывод: Бизнес - направления: производство и сбыт.
- История данных 7-10 лет.

Сначала проектировщик ХД должен в целом оценить ситуацию, касающуюся деятельности организации, для которой он собирается проектировать ХД. Это позволит ему представить проблему в целом и даст определенную свободу в принятии проектных решений.

Допустим, что ХД заказывает организация - Компания, которая занимается производством и продажей телефонов. Компания создана недавно, около 10 лет назад. Сначала около 7 лет она занималась только производством телефонов, сбыт был организован через дистрибьютерскую *сеть*. Спрос на телефоны вырос, и Компания создала свою собственную *сеть* магазинов для продажи телефонов.

Спрос на телефоны продолжал расти, и Компания в прошлом году открыла ряд новых заводов, складов и магазинов. Компания, решила потратить средства на оценку эффективности своего расширения, и в первую *очередь* оценить взаимосвязь между стоимостью и доходом. Компания имеет данные об этих величинах в целом, на уровне заводов, складов и магазинов таких данных было собрано немного. Отчетов, поставляемых ИТ службой Компании, оказалось явно недостаточно для ответа на этот вопрос, и после обсуждения **руководство компании решило создать ХД** для решения этой задачи.

Определение проекта создания ХД: Оценка ситуации в целом

- Компания в прошлом году открыла ряд новых заводов, складов и магазинов.
- Компания, решила потратить средства на оценку эффективности своего расширения, и в первую очередь оценить взаимосвязь между стоимостью и доходом.
- Созданных ИТ службой отчетов оказалось явно недостаточно для ответа на этот вопрос, и после обсуждения **руководство компании решило создать ХД** для решения этой задачи.
- **Вывод:** Очевидно, что имеющихся у Компании данных недостаточно для оценки эффективности.

Спрос на телефоны продолжал расти, и Компания в прошлом году открыла ряд новых заводов, складов и магазинов. Компания, решила потратить средства на оценку эффективности своего расширения, и в первую *очередь* оценить взаимосвязь между стоимостью и доходом. Компания имеет данные об этих величинах в целом, на уровне заводов, складов и магазинов. Таких данных было собрано немного. Отчетов, поставляемых ИТ службой Компании, оказалось явно недостаточно для ответа на этот вопрос, и после обсуждения руководство компании решило создать ХД для решения этой задачи.

Определение проекта создания ХД: Выводы

- Самое главное, **заказчиком является руководство Компании**, не ее конкретная служба.
- Руководство Компании **ожидает** увидеть **информацию**, которая позволит ему **оценивать эффективность** расширения (сворачивания) Компании, и **принимать соответствующие решения**.
- Компания имеет ИТ службу, которая обслуживает OLTP систему, т.е. **источником данных ХД**, вероятнее всего, будет БД этой системы, а модель данных этой БД будет служить отправной точкой разработки ХД

Приведенной выше информации достаточно для того, чтобы проектировщик сделал ряд важных выводов. Во-первых, и это самое главное, заказчиком является руководство Компании, не ее конкретная служба. Во-вторых, руководство Компании ожидает увидеть информацию, которая позволит ему оценивать эффективность расширения (сворачивания) Компании, и **принимать соответствующие решения**. В-третьих, Компания имеет ИТ службу, которая обслуживает *OLTP* систему, т.е. источником данных ХД, вероятнее всего, будет *БД* этой системы, а *модель данных* этой *БД* будет служить отправной точкой разработки ХД.

Определение проекта создания ХД

- **Цель проекта** - создать ХД для обеспечения анализа затрат (стоимости) и прибыли (дохода) для товаров (изделий), произведенных и проданных Компанией.

- **Масштаб проекта** - проект должен быть ограничен учетом прямых затрат и доходов, ассоциированными с продукцией Компании.

После этого *руководитель проекта* разработки ХД должен определить проект. *Определение проекта* должно дать ответ на следующие основные вопросы - что хочет анализировать заказчик, почему это ему нужно, и как он это хочет делать. *Определение проекта* включает *определение* цели и масштаба проекта.

Для нашего примера, цель проекта может быть определена как, создать ХД для обеспечения анализа затрат (стоимости) и прибыли (дохода) для товаров (изделий), произведенных и проданных Компанией.

Масштаб проекта может быть определен как, проект должен быть ограничен учетом прямых затрат и доходов, ассоциированными с продукцией Компании.

Теперь проектировщик ХД может приступить к решению задачи – сбор требований к хранилищу данных.

Сбор требований: направления бизнеса

- Допустим, что следующие важные направления деятельности Компании должны найти отражения в ХД:
 1. Жизненный цикл производства;
 2. Структура продаж;
 3. Структура организации;
 4. Определение затрат и прибыли;
 5. Что хочет делать потенциальный пользователь в ХД.

Для реализации проекта создается *команда*, которая включает *специалистов предметной области* и ИТ службы. Допустим, что эта *команда* определила следующие важные направления деятельности Компании, которые должны найти отражения в ХД:

Жизненный цикл производства

Структура продаж

Структура организации

Определение затрат и прибыли

Что хочет делать потенциальный *пользователь* в ХД.

Общая схема работы по сбору требований

- Примеры основных вопросов, на которые должен быть получен ответ, следующие:
 - Данные о ком или о чем (люди, группы, организации) представляют интерес для пользователя ХД?
 - Какие функции используются для анализа данных пользователем?
 - Почему пользователю необходимы такие данные?
 - Когда (в какие моменты времени) такой анализ должен быть выполнен пользователем?
 - Где (организационно) будет выполняться анализ?
 - Какие показатели производительности или функции состояния будут анализироваться?

Требования, определенные в этой точке жизненного *цикла*, используются для построения модели ХД.

Вернемся к нашему примеру о разработке ХД Компании.

Сбор требований: Вопросы

- Методы получения бизнес – требований могут быть разбиты на две категории:
 1. Управляемые пользователем. Основывается на определении требований исходя из функций, выполняемых пользователями.
 2. Управляемые источниками данных. Основывается на определении требования посредством использования имеющихся данных в оперативных системах.

Преимуществом первого является то, что бизнес- *аналитик* фокусируется на потребностях пользователя. При таком подходе исследуется меньший объем данных, он лучше описывает требования к ХД и может быть выполнен быстрее.

Преимуществом второго подхода является то, бизнес-*аналитик* исходит из имеющихся данных и пытается построить на их основе показатели для ХД. Этот подход предполагает исследование ER модели Компании, реализованной в *OLTP* системах, требует значительно больше времени, чем первый и требует нескольких итераций для приведения требований в согласии с пользователями.

Мы в нашем примере остановимся на использовании первого метода.

Сбор требований: Жизненный цикл производства

- Каждый завод имеет группу, которая разрабатывает процесс производства нового продукта. Когда новый продукт получает одобрение, информация о нем заносится в номенклатуру продукции компании. После этого все заводы могут выпускать продукт.
- Продукт имеет базовый набор комплектующих компонент. Дополнительные комплектующие компоненты используются для создания специфической модели продукта.
- Компания производит 200 моделей. Политика компании строится таким образом, что число выпускаемых моделей остается постоянным. Это означает, что количество новых моделей приблизительно равно количеству моделей, снятых с производства.

Каждый завод имеет группу, которая анализирует идею нового продукта. Только после того, как процесс производства полностью определен, и одобрение нового продукта получены, *информация* о продукте добавляется в номенклатуру продукции компании. После этого все заводы могут выпускать продукт.

Продукт имеет базовый набор комплектующих *компонент*. Дополнительные комплектующие компоненты используются для создания специфической модели продукта.

Пусть, Компания производит 200 моделей. Политика компании строится таким образом, что число выпускаемых моделей остается постоянным. Это означает, что количество новых моделей приблизительно равно количеству моделей, снятых с производства.

- Примерно для 10 моделей в неделю проверяется изменение затрат и прибыли. Для каждой модели каждого продукта принимается решение, давать или не давать скидки на данную модель.
- Когда модель является приемлемой для назначения скидки, продавцы могут давать скидки клиентам, если покупатель приобретает большую партию продукции этой модели или их комбинации. Но заведующий складом розничной продажи должен одобрить такую скидку.

Примерно для 10 моделей в неделю проверяется изменение затрат и прибыли. Для каждой модели каждого продукта принимается решение, давать или не давать скидки на данную модель. Когда модель является приемлемой для назначения скидки, продавцы могут давать скидки клиентам, если *покупатель* приобретает

большую партию продукции этой модели или их комбинации. Но заведующий складом розничной продажи должен одобрить такую скидку.

- Каждый завод управляет запасом продукции данной модели. Когда остаток продукции данной модели становится меньше определенного уровня, создается заказ для производства продукции данной модели.
- После того, как продукция данной модели будет произведена, она остается на заводе до тех пор, она не будет затребована отделом сбыта. Отдел сбыта дает показатель продаваемости модели.
- Когда принято решение приостановить производство данной модели, через 6 месяцев после продажи последней модели данные удаляются из БД.

Каждый завод управляет запасом продукции данной модели. Когда *остаток (quantity on hand)* продукции данной модели становится меньше определенного уровня, заказ на работу создается для производства продукции данной модели. После того, как продукция данной модели будет произведена, она остается на заводе до тех пор, она не будет затребована отделом сбыта (*sales outlet*). Отдел сбыта дает показатель продаваемости модели.

Когда принято решение приостановить производство данной модели, данные о ней хранятся в БД организации в течение 6 месяцев после того, как последняя *единица* продукции данной модели будет продана или придет в негодность.

Данные о продукции удаляются в тот момент, когда удаляются данные о последней модели этой продукции.

Сбор требований: Структура продаж

- Существуют два типа отделов сбыта - отдел корпоративных продаж и отдел розничной продажи. Отдел корпоративных продаж продает только оптовым покупателям.
- Оптовый покупатель может предоставлять счет либо непосредственно в отдел корпоративных продаж, либо по факсу. По счету продукция отгружается прямо с завода. Покупатель может иметь несколько пунктов отгрузки. Покупатель может размещать счета в различных отделах продаж.
- Оптовый покупатель определяется 30 продажами. Организация в настоящий момент обслуживает 3000 оптовых покупателей.
- Отдел корпоративных продаж отправляет документы в пункты приема оптовых покупателей. Если покупатель имеет несколько пунктов приема товара, то отдел корпоративных продаж направляет соответствующие документы в каждый.
- Отдел корпоративных продаж создает в среднем 500 счетов в день, 5 дней в неделю. Каждый счет включает в среднем 10 моделей продукции.

Существуют два типа отделов сбыта - отдел корпоративных продаж (*corporate sales office*) и отдел розничной продажи (*retail store*). Отдел корпоративных продаж продает только оптовым покупателям. Оптовые покупатели определяются *по* закупочной цене (*wholesale price*), независимо от предоставленных скидок. Оптовый *покупатель* определяется 30 продажами. Организация в настоящий момент обслуживает 3000 оптовых покупателей.

Оптовый *покупатель* может предоставлять счет либо непосредственно в отдел корпоративных продаж, либо *по* факсу. Эти счета отгружаются прямо с завода. *Покупатель* может иметь несколько пунктов отгрузки. *Покупатель* может размещать счета в различных отделах продаж.

Отдел корпоративных продаж отправляет документы в пункты приема оптовых покупателей. Если *покупатель* имеет несколько пунктов приема товара, то отдел корпоративных продаж направляет соответствующие документы в каждый. Отдел корпоративных продаж создает в среднем 500 счетов в день, 5 дней в неделю. Каждый счет включает в среднем 10 моделей продукции.

Сбор требований: Структура продаж

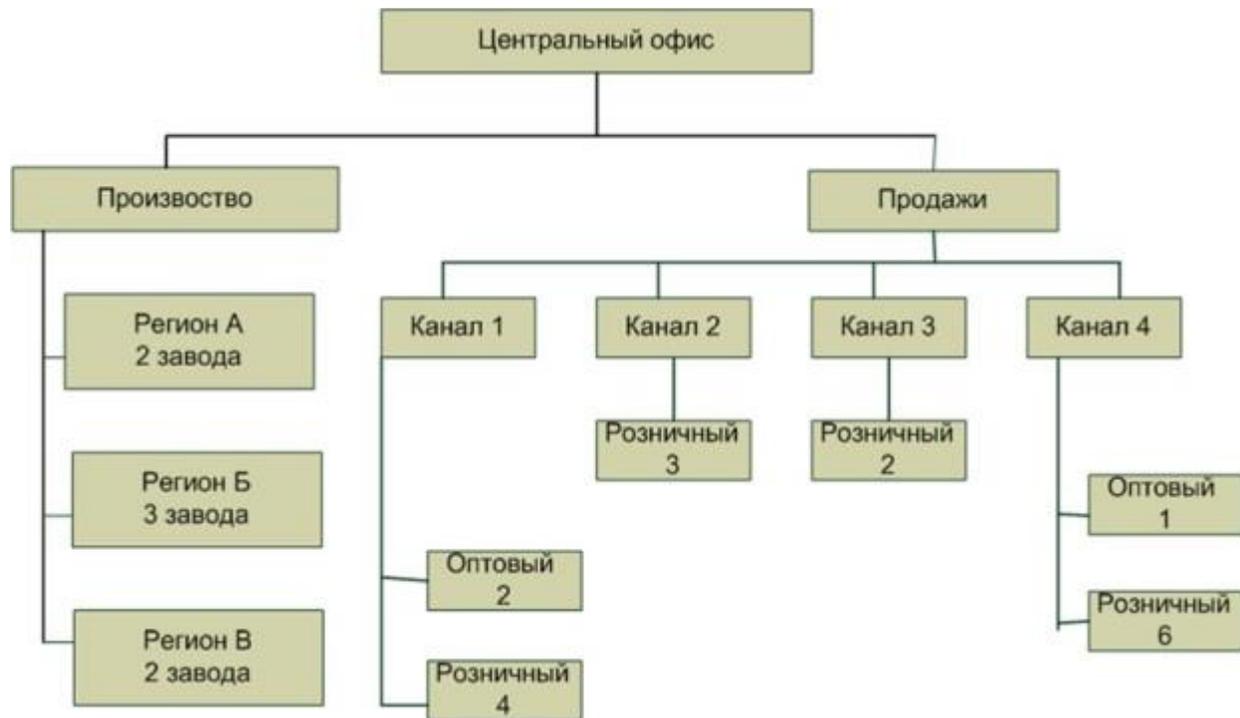
- Отдел розничной продажи продает за наличный расчет. Не зависимо от предоставления скидок, цена товара меняется.
- Хотя на каждую продажу продукции оформляется счет, организация не ведет учет покупателей для розничной продажи.
- Отдел розничной продажи генерирует в среднем 1000 счетов в день, семь дней в неделю. Каждый счет содержит оплату в среднем двух моделей продукции.
- Каждый отдел связан только с одним заводом.

Склад розничной продажи продает за наличный расчет. Не зависимо от предоставления скидок, *цена товара* меняется. Хотя на каждую продажу продукции оформляется счет, организация не ведет учет покупателей для розничной продажи.

Каждый склад связан только с одним заводом. Заведующий складом отвечает за то, какая продукция хранится и продается с его склада.

Склад розничной продажи генерирует в среднем 1000 счетов в день, семь дней в неделю. Каждый счет содержит оплату в среднем двух моделей продукции.

Сбор требований: Структура организации



Сбор требований: Определение доходов и расходов

- Потенциальные пользователи системы анализируют показатели расхода и дохода. Стоимость модели определяется как сумма стоимостей всех компонент модели. Для каждой модели установленная цена единицы товара умножается на количество продаж. Суммирование результатов по всем позициям заказа есть доход от продажи данной модели.
- Рассмотрим возможную проблему в OLTP системе.

Потенциальные пользователи системы анализируют показатели расхода и дохода. Для каждой модели продукции *стоимость* каждого компонента умножается на число *компонент*, используемых при производстве модели. *Стоимость* модели определяется как сумма стоимостей всех *компонент* модели. Для каждой модели установленная цена единицы товара (*negotiated unit selling price*) умножается на количество продаж. Суммирование результатов *по* всем позициям заказа (*order lines*) есть *прибыль* (*revenue*) от продажи данной модели.

При попытке связать *стоимость* модели с прибылью от ее продажи, обнаружилось, что каждая модель производилась и добавлялась к остатку в запасе, *стоимость* этой единицы продукции модели не могла быть определенно идентифицирована. Даже, если *стоимость компонент* контролировалась, она использовалась только для вычисления текущего значения запаса (*inventory*). *Фактическая стоимость* записывалась только в финансовой системе организации, без какой либо ссылки на количество произведенной продукции.

Результатом такого определения было следующее. 1) В *OLTP* системе должна быть изменена процедура определения стоимости произведенной модели. Поскольку *стоимость* компоненты меняется часто и на незначительную величину, то *прибыль* от продажи модели всегда заносится *по* текущей стоимости единицы данной модели, независимо от фактических затрат на производство этой модели.

Сбор требований: 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?

В процессе работы над проектом одним из первых действий команды разработчиков было *определение* набора типовых запросов, на который пользователи хотели бы получить ответы в результате создания хранилища данных. Был определен следующий *список* главных вопросов:

Какова величина среднего остатка продукции на складе и уровень запасов, при котором подается заказ (*reorder level*) за текущий месяц для каждой модели на каждом заводе.

Какова величина суммарных затрат (*total cost*) и суммарной прибыли (*revenue*) *по* каждой модели, проданной сегодня, и просуммированной *по* отделу сбыта (*outlet*), типу отдела сбыта (*outlet type*), области (*region*) и *по* структуре продаж Компании (*corporate sales levels*)?

Какова величина суммарных затрат (*total cost*) и суммарной прибыли (*revenue*) для каждой модели, проданной сегодня, и просуммированной *по* заводам (*manufacturing plant*) и *по* областям (*region*)?

Какой *процент* моделей получили скидки, и какие из них были проданы *по факту* со скидкой (в процентах) *по складам (store)* для всех продаж (*all/sales*) на этой неделе? В этом месяце?

Для каждой модели, проданной в текущем месяце, какой был *процент* продаж с розничной торговли, с оптовой торговли *по безналичному расчету (order desk)*, с оптовой торговли через продавцов (*salesperson*)?

Какие модели и какого типа продукция не продавалась в течение последнего месяца? В течение последней недели?

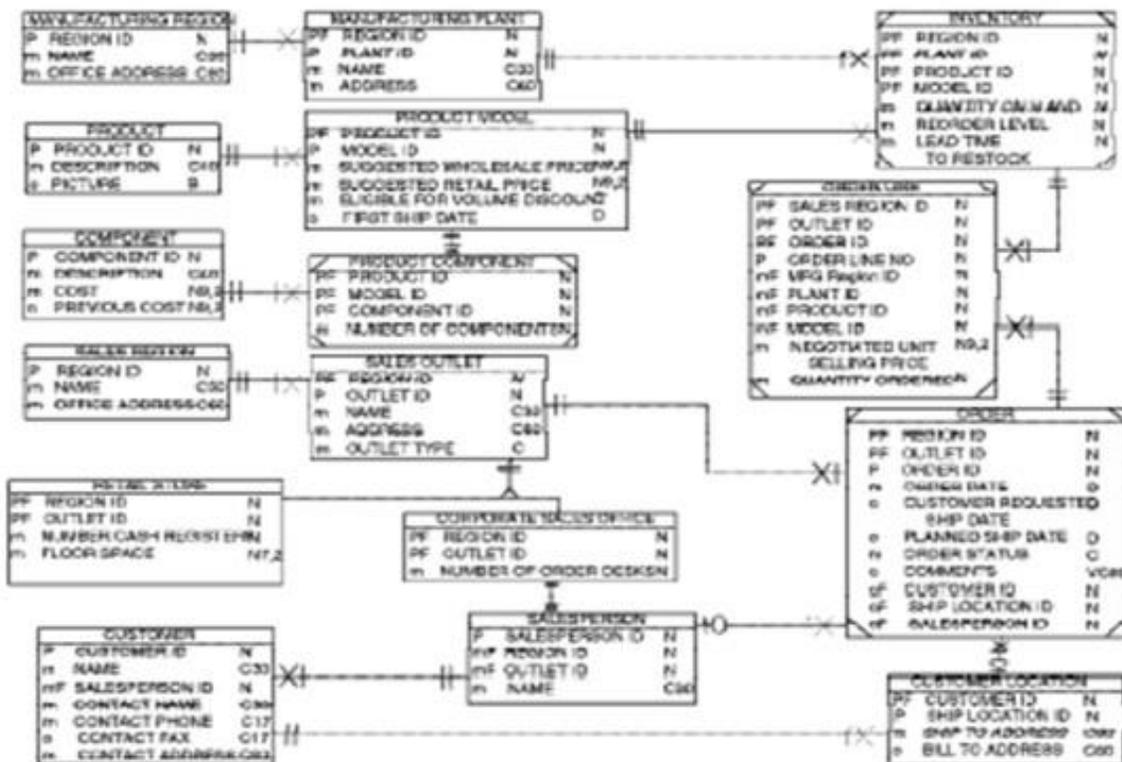
Какие пять моделей, проданных за последний месяц, принесли наибольшую *прибыль (total revenue)*? *По* продажам за квартал (*quantity sold*)? *По* суммарным затратам?

Какие отделы сбыта (*sales outlets*) не имели продаж в течение последнего месяца для каждой модели в каждом из пяти топ-списков?

Какие продавцы не имели ни одной записи о продажах за последний месяц для каждой модели в каждом из трех списков 5 моделей?

Помимо прочего, пользователи хотят получать ответы на аналогичные вопросы за последние пять лет и в будущем.

Сбор данных: Корпоративная ER модель



Корпоративная ER модель *OLTP* Компании. Бизнес – *аналитик* и проектировщик ХД должны изучить эту модель для того, чтобы выяснить есть необходимые данные для удовлетворения бизнес – требований пользователей.

Структура записи по товарам

Имя	Тип данных	Длина	
ProductID	Numeric	5	
ModelID	Numeric	5	
Product Descr	Character	40	
Suggested Wholesale Price	Numeric (9,2)	5	Оптовая цена
Suggested Retail Price	Numeric (9,2)	5	Розничная цена
Eligible for Volume Discount	Character	1	Скидка при оптовой продаже

После определения бизнес - требований необходимо найти данные, необходимые для построения ХД, которые эти бизнес – требования удовлетворили.

Бизнес – *аналитик* изучает ER модель *OLTP* системы Компании, чтобы найти в ее *БД* объекты, содержание необходимую информацию.

Структура записи для товаров и моделей может иметь вид, как на слайде.

Структура записи по комплектующим

Имя	Тип данных	Длина	
ComponentID	Numeric	5	
ProductID	Numeric	5	

ModelID	Numeric	5	
Component Description	Character	40	
Unit Cost	Numeric (9,2)	5	Цена на ед. продукции
Number of Components	Numeric	5	

После определения бизнес - требований необходимо найти данные, необходимые для построения ХД, которые эти бизнес – требования удовлетворили.

Бизнес – *аналитик* изучает ER модель *OLTP* системы Компании, чтобы найти в ее *БД* объекты, содержание необходимую информацию.

Структура записи для товаров и моделей может иметь вид, как на слайде.

2. 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

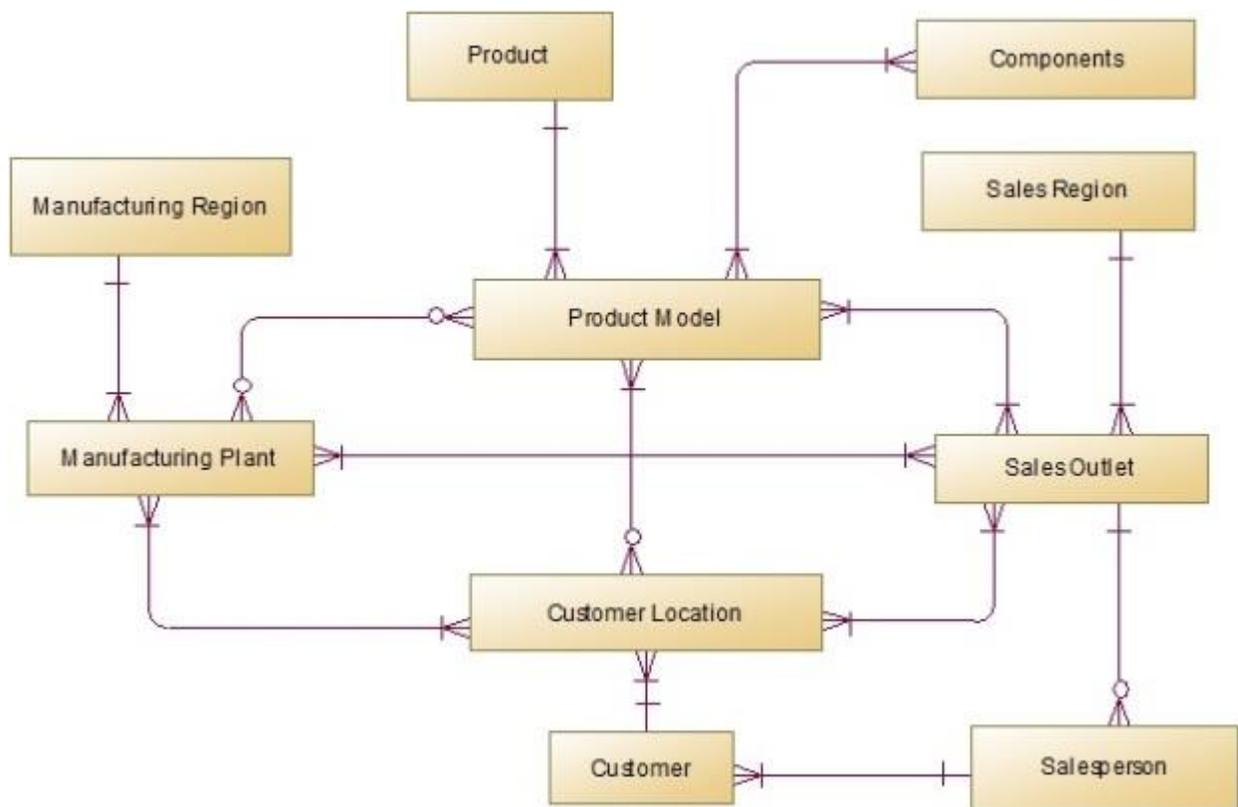
- Определение кандидатов в измерения, их атрибутов и *потенциальных ключей*.
- Определение кандидатов в показатели (метрики).
- Определение кандидатов в факты из бизнес – требований.
- Определения кандидатов в факты исходя из семантического анализа предметной области.
- Документирование проделанной работы.

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



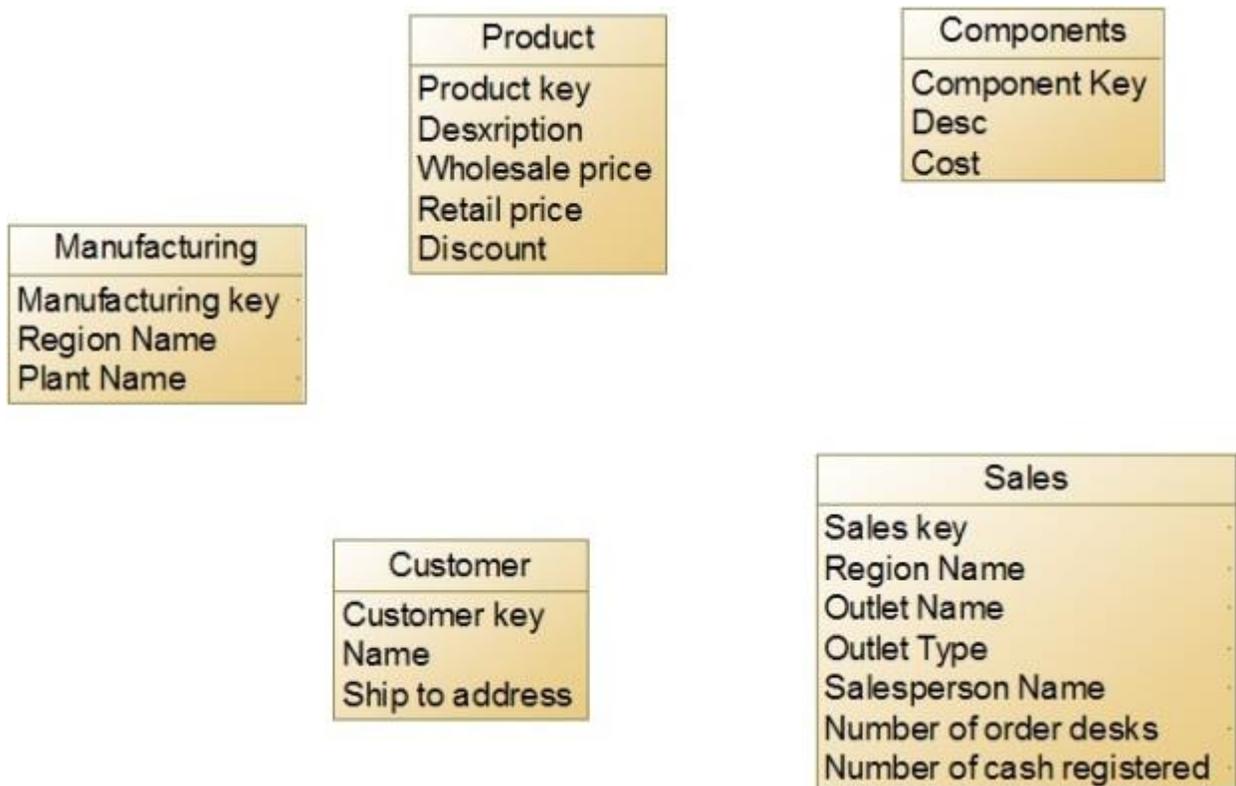
В качестве примера мы будем использовать ER модель, которую рассматривали на прошлом семинарском занятии.

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 складе (Average *Quantity on Hand*)
 - Определить Суммарная прибыль (Total Revenue),
- то кандидатами на показатели могут быть
 - Средний остаток продукции на складе (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.

Измерения и показатели

- Измерения требуются для интерпретации показателей.
- Например, интерпретация такого показателя как остаток на складе требует знания, к какому виду продукции этот остаток относится, где этот остаток размещается (на складе какого завода) и период времени, с которым значение этого показателя связывается.
- Такая постановка вопроса приводит в определению следующих измерений для этого показателя: Продукт (Product), Завод (Manufacturing) и Время (Time).
- Если в качестве показателя взять показатель Суммарная прибыль (Total Revenue), анализируемый в запросе 2, то для него следует определить такие измерения как Продажи (Sales), Продукт (Product) и Время (Time).
- Для показателя запроса 3 Суммарные затраты (Total Cost) в качестве измерений следует взять Продукт (Product), Завод (Manufacturing) и Время (Time).

После определения уровня структуризации данных следует уточнить *список* кандидатов в измерения.

Таким образом, измерения являются координатами многомерного пространства изменения показателей, в которых последние получают свою интерпретацию в рамках *предметной области*. Как видно из выше сказанного, методом определения кандидатов в измерения является *анализ* контекста запросов к ХД.

Измерения, показатели и запросы

	Sales	Manufacturing	Product
ProductID	Numeric 5		
1		X	X
2	X		X
3		X	X
4	X		
5	X		X
6			X
7			X
8	X		X
9	X		X

На следующем шаге проектировщик должен объединить всех претендентов на измерения и исследовать их с точки зрения всех известных запросов к ХД. Для нашего базового учебного примера проектировщик может идентифицировать четыре измерения: Продажи (Sales), Продукт (Product), Завод (*Manufacturing*) и Время (*Time*).

Хорошим элементом анализа при построении модели в рамках нашего базового учебного примера является построение матрицы соответствия бизнес-требований, измерений и показателей. Бизнес-требования определяются списком главных вопросов. Таким образом, имеем первую итерацию матрицы:

Запросы и показатели

	1	2	3	4	5	6	7	8	9
average <i>quantity</i> on hand	X								
Total Cost		X	X				X		
Total Revenue		X					X		
<i>Percentage</i> of models eligible for discount				X			X		
<i>Percentage</i> of models eligible for discount				X					
<i>Percentage</i> of models sold through retail <i>outlet</i>					X				
<i>Percentage</i> of models sold through sales office order desk					X				
<i>Percentage</i> of models sold through sales office salesperson					X				

Средний *остаток* продукции на складе (*average quantity on hand*)

Суммарные *затраты* (*Total Cost*)

Суммарная *прибыль* (*Total Revenue*)

Процент моделей со скидкой (*percentage of models eligible for discount*)

Процент моделей со скидкой (*percentage of models eligible for discount*) проданных по факту

Процент моделей проданных через розничную торговлю (*percentage of models sold through retail outlet*)

Процент моделей проданных через оптовую. торговлю (*percentage of models sold through sales office order desk*) по безналичному расчету

Процент моделей проданных через оптовую. торговлю (*percentage of models sold through sales office salesperson*) продавцами

Промежуточный итог

- Более естественным подходом к идентификации измерений является исследование свойств потенциальных показателей в контексте бизнес-процессов и бизнес-правил. Определение потенциальных измерений в контексте бизнес - требований с этой точки зрения является первым этапом решения задачи выбора измерений.
- Например, показатель *Прибыль с продажи* связан с операцией продажи продукции, которая вовлекает в бизнес-процесс покупателя этой продукции, и, естественным образом приводит к еще одному измерению Покупатель (Customer).
- К решению задачи выбора измерений мы вернемся позже, после обсуждения решения других задач, которые непосредственно повлияют на окончательную картину многомерного пространства.

Определение кандидатов в факты

- Для извлечения фактов следует снова обратиться к бизнес -требованиям.
- Рассмотрим вопрос 1. Анализ этого запроса позволяет выделить такой факт как Запас (Inventory). В этом запросе вычисляется средний остаток продукции на складе заданного завода для данной модели в течение заданного периода времени. Факт Запас интерпретируется в трех измерениях: Время, Продукт и Завод.
- Рассмотрим вопрос 2 и 3 нашего учебного примера. Оба вопроса связаны с фактом Продажи (Sales). Этот факт объединяет два параметра - Суммарные затраты и Суммарная прибыль, которые являются независимыми параметрами и интерпретируются в трех измерениях: Время, Продукт и Завод.

Поскольку параметры вместе с измерениями определяют факты, то после предварительного определения параметров и измерений можно перейти к определению кандидатов в факты.

Классификация фактов

- Факты могут генерироваться бизнес-операциями или событиями хозяйственной деятельности организации. Так, например, с фактом Продажа связаны следующие операции и события: что продано, кто продал, где и когда было продано, как много было продано, какие скидки были использованы и т.д.
- Факты могут представлять собой состояния объектов предметной области (сферы деятельности организации). Например, состояние запаса - это ответы на вопросы: что храниться, где хранится, как много хранилось в течение данного периода и т.д.
- Факты могут также представлять изменение состояния объектов хозяйственной деятельности организации (предметной области ХД). Изменение запаса представляет собой перемещение определенного количества запаса из одного места хранения в другой или во внешнюю организацию.

Факты являются центральным элементом *многомерной модели*. Проектировщику ХД очень важно правильно выбрать факты, соответствующие задачам анализа данной *предметной области*. Однако выбор фактов на основе предварительно определенных параметров и измерений носит технологический характер.

Факты следует определять исходя из семантики *предметной области* и перспектив деятельности организации (стратегического бизнес-плана организации). При выборе фактов с такой позиции проектировщик ХД может исходить из следующей их классификации:

Правила отбора фактов

1. Каждый факт должен существовать с точки зрения хозяйственной деятельности организации.
2. Связанные с хозяйственной деятельностью организации факты генерируются в результате операций хозяйственной деятельности организации, событий хозяйственной деятельности организации, объектами хозяйственной деятельности, чье состояние интересно с точки зрения анализа, объектами хозяйственной деятельности, изменение состояние которых интересно с точки зрения анализа
3. Каждый факт должен иметь уникальную идентификацию в сфере хозяйственной деятельности организации.

4. Уровень структуризации основных измерений каждого факта

Проектировщик ХД должен уметь выбирать связанные с хозяйственной деятельностью организации факты, чтобы затем построить *многомерную модель*. Отбирая факты для построения *многомерной модели* ХД, проектировщику следует придерживаться следующих правил:

Выбор фактов на основе семантики ПрО

- Продажи являются примером факта, представляющего операцию хозяйственной деятельности или событие хозяйственной деятельности организации.
- Если руководство хочет оценить продуктивность продаж, то оно использует показатели Суммарные затраты или Суммарная прибыль, которые связаны с событием продажи товара. Факт продажи представляет "что-то, что произошло".
- Событие или операция хозяйственной деятельности может принадлежать как внешней, так и самой организации.

Проектировщик ХД должен уметь выбирать связанные с хозяйственной деятельностью организации факты, чтобы затем построить *многомерную модель*. Отбирая факты для построения *многомерной модели* ХД, проектировщику следует придерживаться следующих правил:

Выбор фактов на основе семантики ПрО

- Поиск фактов для ХД в рамках хозяйственной деятельности организации приводит к идентификации дополнительных показателей.
- Например, если факт Продажи определяется через призму операции продажи товара, то Количество (*Quantity Sold*) товара проданного в результате операции следует добавить к показателям, которые нужно сохранять в ХД. Количество проданного товара проходит через ряд связанных бизнес - операций, начиная с оформления счета покупателя до получения подтверждения о получении товара покупателем.

Выбор фактов на основе семантики ПрО

- Рассмотрим факт Продажа. Операция хозяйственной деятельности предполагает, что произошло изменение в предметной области. Обычно руководители хотят знать, какой эффект произошел в результате этих изменений, и, следовательно, этот эффект нужно как-то оценить. Поэтому, факт, который генерируется в результате операции хозяйственной деятельности, непосредственно связан с некоторой мерой эффекта этой деятельности, т.е. некоторым числовым показателем.

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

Выбор фактов на основе семантики ПрО

- Если руководителей интересует, когда произошла продажа, то факт продажи интерпретируется как событие хозяйственной деятельности. Обычно, для фактов, генерируемых в результате событий хозяйственной деятельности, не ставят в соответствие какой либо показатель, который измеряет эффект от этого события. Факт Продажа означает, что в некоторый момент времени определенный продукт с определенного склада и определенного завода был продан. Факты, связанные с событиями, называются не измеряемыми фактами (*factless facts* или *measureless facts*).

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

Выбор фактов на основе семантики ПрО

- Факт Уровень запаса (Inventory) представляет состояние объекта хозяйственной деятельности - запасом товара на определенном заводе в заданный период времени. Состояние такого объекта как запас можно оценить с помощью показателя Средний остаток продукции. Заметим, что в случае фактов отражающих состояние измерение Время интерпретируются в терминах период времени, а показатель, измеряющий значение состояния объекта, представляет собой численное значение в течение этого периода времени.

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

Выбор фактов на основе семантики Про

- Одна из основных проблем для фактов, представляющих изменение состояние, - это выбор адекватного показателя. Например, если для заданного периода времени частота смены состояния достаточно высока, то показатель Средний остаток продукции становится для факта Уровень запаса не слишком подходящим, поскольку перестает отражать динамику изменения запаса.
- При решении этой проблемы можно поступать несколькими способами. Оставляя уровень структуризации измерения времени тем же самым, можно выбрать другой показатель, например, Остаток продукции на складе. Можно пойти по пути увеличения уровня структурированности представления факта Уровень запаса, сохраняя в ХД значение показателя не за месяц, а за день. Принимая такое решение, проектировщик ХД должен задать себе вопрос - позволяет ли источник данных получать значения показателя на таком уровне структуризации. Можно посмотреть на проблему совсем с другой стороны, сохранять в ХД не состояние запаса, а изменение состояния запаса. Последний показатель будет лучше отражать динамику управления запасом.

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

Выбор фактов на основе семантики Про

- Рассмотрим показатель Средний остаток товара для факта Уровень запаса продукции. Пусть уровень структуризации измерения Время увеличится до одного дня, но частота изменения состояния объекта по-прежнему останется высокой. С этой точки зрения проектировщик должен увеличить уровень структуризации измерения Время до часа, чтобы ХД правильно отражало динамику изменения запаса.
- Обратим внимание, что в такой постановке задачи аналитиков интересует анализ поведения объекта хозяйственной деятельности, в данном случае, изменение запаса продукции. Тогда можно рассмотреть факт Уровень запаса с иной точки зрения, т.е. изменить его интерпретацию: отслеживать изменения состояния объекта и сохранять в модели показатель Количество перемещенного запаса (*Quantity Moved*).
- Окончательную точку в решении проблемы выбора показателей фактов, описывающих состояние или изменение состояния объекта хозяйственной деятельности проектировщик ХД может решить, либо обращаясь к потенциальным пользователям (с чем предпочитают работать пользователи - с состояниями объектов или изменениями состояния), либо опираясь на методы моделирования времени в ХД (о чем пойдет речь далее).

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

Выводы

- Проектировщик ХД после анализа требований должен иметь следующие документы:
 - Перечень кандидатов в измерения и их анализ;
 - Перечень кандидатов в показатели (метрики) и их анализ;
 - Перечень кандидатов в факты, их классификацию и анализ.

В чем состоит различие между фактами, генерируемыми операциями хозяйственной деятельности, и событиями хозяйственной деятельности.

3. На основе определенных в процессе анализа требований кандидатов в измерения, показатели и факты построить эскиз *многомерной модели* ХД.

- Задачи:
 - Измерения;
 - Показатели;
 - Факты;
 - Пересмотр показателей и измерений (агрегация и гранулированность);
 - Объединение фактов.

Зачем нужно анализировать бизнес-требования

- Во-первых, это все равно придется делать, поскольку одной из главных целей проекта является удовлетворение таких требований.
- Во-вторых, очень важно, когда это делать и как. Если серьезный анализ требований отнести на стадию, скажем разработки приложений, то может возникнуть ряд обратных задач проектирования, решение которых может привести к изменению структуры модели.
- Например, показатель Средний остаток может быть заменен на Количество перемещенного запаса, что повлечет внесение изменений в результаты всех последующих стадий проектирования и разработки ХД.

Построение эскиза многомерной модели ХД

- Эта задача решается в пять этапов:
 - Определение измерений;
 - Определение показателей (агрегация и *гранулированность*);
 - Определение фактов;
 - Анализ фактов, показателей и измерений.
 - Построение эскиза *многомерной модели*.

Определив потенциальные измерения, показатели и факты, проектировщик ХД на следующем этапе может приступить к созданию эскиза *многомерной модели* данных.

Измерения

- На предыдущем семинаре мы определили кандидаты в измерения.
- На этапе создания эскиза *многомерной модели* проектировщик ХД должен решить, какие измерения должны быть в модели.
- Учитывая тот факт, что ХД, как правило создается для хранения и анализа исторических данных о хозяйственной деятельности организации, необходимо добавить в модель измерение "Время".

Измерения модели ХД

TIME		
<u>Time_Key</u>	bigint	<pk>
Date	date	
Week	integer	
Mouth	char(10)	

PRODUCT		
<u>Product_Key</u>	bigint	<pk>
Descritoin	char(40)	
Model_Code	char(10)	
Unit Cost	numeric(9,2)	
Wholesale Price	numeric(9,2)	
Retail Price	numeric(9,2)	
Column_7	<Undefined>	
Effective from Date	date	
Effective to Date	date	

SALES		
<u>Sales_Key</u>	bigint	<pk>
Region Name	char(30)	
Ourlet Name	char(30)	
Outlet Type	char(10)	
Salesperson Name	char(30)	

MANUFACTURING		
<u>Manufacturing_Key</u>	bigint	<pk>
Region Name	char(30)	
Plant Name	char(30)	

Факты

- Факты строятся на основе выбранных измерений и показателей (метрик).
- Построение фактов есть процесс группировки измерений и показателей иаким образом, чтобы можно было удовлетворить бизнес – требованиям.
- Первоначально проектировщик ХД фожет создать один факт для всех запросов.
- Для каждого набора показателей точно связанных с одними и теми же измерениями создадим отдельный факт.

Заметим, что вопросы 6, 8 и 9 не имеют показателей, связанных с ними. Не объединить ли *запрос* 6 с запросами 5 и 7 в факт 4, а запросы 8 и 9 с запросом 2 в факт 2, тем самым избавившись от фактов без показателей. (такие факты называются *factless facts*, потому что они фиксируют событие: продажу товара в определенный момент времени (факты 2 и 3) в определенном месте (факт 2 только). Никакие показатели для них не требуются.

Факты модели ХД: запросы 1 и 3

Fact 1		
Time_Key	bigint	<fk1>
Product_Key	bigint	<fk2>
Manufacturing_Key	bigint	<fk3>
average quantity on hand	numeric(6,2)	
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Reorder Level	numeric	

Факты модели ХД: запросы 2,8 и 9

FACT 2		
Sales_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
Product_Key	bigint	<fk3>
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	

Факты модели ХД: запросы 5,6 и 7

FACT 3		
Product_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Total Quantity Sold	numeric	
percentage of models sold through retail outlet	numeric(5,2)	
percentage of models sold through sales office order desk	numeric(5,2)	
percentage of models sold through sales office salesperson	numeric(5,2)	

Факты модели ХД: запрос 4

FACT 4		
Sales_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
percentage of models eligible for discount	numeric(5,2)	
percentage of models eligible for discount actually discounted	numeric(5,2)	

Гранулированность, аддитивность и объединение фактов

- Гранулированность есть уровень детализации факта, с которой он будет храниться в ХД.
- Аддитивность – это возможность суммировать показатели в ХД. Тесно связана с *гранулированностью*.
- Факты бывают аддитивные, неаддитивные и полуаддитивные.

Аддитивность фактов

- Факты бывают аддитивные, неаддитивные и полуаддитивные:
 - Проценты – неаддитивные факты (бессмысленно суммировать);
 - Баланс - полуаддитивный факт (можно складывать только в определенный момент времени, но не в различные);
 - Доход – *аддитивный факт* (можно суммировать по всем измерениям).

Гранулированность и аддитивность фактов

- Рассмотрим ФАКТ 1:
 - average quantity on hand определяется ежемесячно;
 - total cost и total revenue определяются ежедневно.
- Следовательно, либо факт нужно разбить на два факта, либо изменить измерение Time (ввести уровень гранулированности – день).

average quantity on hand – неаддитивный показатель, необходимо хранить *actual quantity on hand* и позже в запросе вычислять среднее. Лучше сделать этот показатель аддитивным.

Гранулированность и аддитивность фактов

- Рассмотрим ФАКТ 2: он имеет проблему с измерением TIME для запроса 2 (ежедневно) и запросов 8 и 9 (ежемесячно):
- Следовательно, нужно изменить измерение Time (ввести уровень гранулированности – день). Это также позволит суммировать по месяцам.

average quantity on hand – неаддитивный показатель, необходимо хранить *actual quantity on hand* и позже в запросе вычислять среднее. Лучше сделать этот показатель аддитивным.

Гранулированность и аддитивность фактов

- Рассмотрите факты ФАКТ 3 и 4 самостоятельно.

average quantity on hand – неаддитивный показатель, необходимо хранить *actual quantity on hand* и позже в запросе вычислять среднее. Лучше сделать этот показатель аддитивным. Целесообразно все показатели сделать аддитивными.

Факты. Промежуточный итог

Fact 1		
Time_Key	bigint	<fk1>
Product_Key	bigint	<fk2>
Manufacturing_Key	bigint	<fk3>
Quantity on hand	numeric	
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Reorder Level	numeric	

FACT 2		
Sales_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
Product_Key	bigint	<fk3>
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	

FACT 4		
Sales_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
Number of models eligible for discount	numeric	
Quantity of models eligible for discount actually sold	numeric	
Quantity of models sold at a discount	numeric	

FACT 3		
Product_Key	bigint	<fk1>
Time_Key	bigint	<fk2>
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Total Quantity Sold	numeric	
Quantity of models sold through retail outlet	numeric	
Quantity of models sold through sales office order desk	numeric	
Quantity of models sold through sales office salesperson	numeric	

FACT 1

Quantity on hand

Time

FACT 2

Time

FACT 3

percentage of models sold through retail outlet на *Quantity of models sold through retail outlet*

percentage of models sold through sales office order desk на *Quantity of models sold through sales office order desk*

percentage of models sold through sales office salesperson на *Quantity of models sold through sales office salesperson*

Time

ФАКТ 4

percentage of models eligible for discount на Number of models eligible for discount

percentage of models eligible for discount actually discounted на

Quantity of models eligible for discount actually *bold*

Quantity of models *bond* at a discount

Объединение фактов

- ФАКТ 1 можно оставить без изменения (почему?).
- У ФАКТ 2 те же измерения, что у 3 и 4. Можно исследовать возможность объединить эти факты.
- ФАКТ 2 и ФАКТ 3 можно объединить.
- Новый ФАКТ 2 можно объединить с ФАКТ 4.
- Добавим еще одно измерение "Продавец" (Seller).

Рассмотрим *fact* 1: ясно, что *total* и *and total* revenue не может быть испорчен измерением sales. Однако это не так для *quantity on hand* или *reorder level*. Фактически *quantity on hand* не следует трогать, поскольку с измерениями *product* и *manufacturing* для него все нормально. Не будем изменять этот факт.

Fact 2 можно объединить с *Fact* 3, добавляя измерение sales к *Fact* 2.

Объединение фактов

ФАКТ 2		
Sales_Key	bigint	<f
Time_Key	bigint	<f
Product_Key	bigint	<f
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Total Quantity Sold	numeric	

ФАКТ 2		
Sales_Key	bigint	<f
Time_Key	bigint	<f
Product_Key	bigint	<f
Total Cost	numeric(9,2)	
Total Revenue	numeric(9,2)	
Total Quantity Sold	numeric	
Discount Amount	numeric(9,2)	

Рассмотрим *fact* 1: ясно, что *total* и *and total* revenue не может быть испорчен измерением sales. Однако это не так для *quantity on hand* или *reorder level*. Фактически *quantity on hand* не следует трогать, поскольку с измерениями *product* и *manufacturing* для него все нормально. Не будем изменять этот факт.

Fact 2 можно объединить с *Fact* 3, добавляя измерение sales к *Fact* 2.

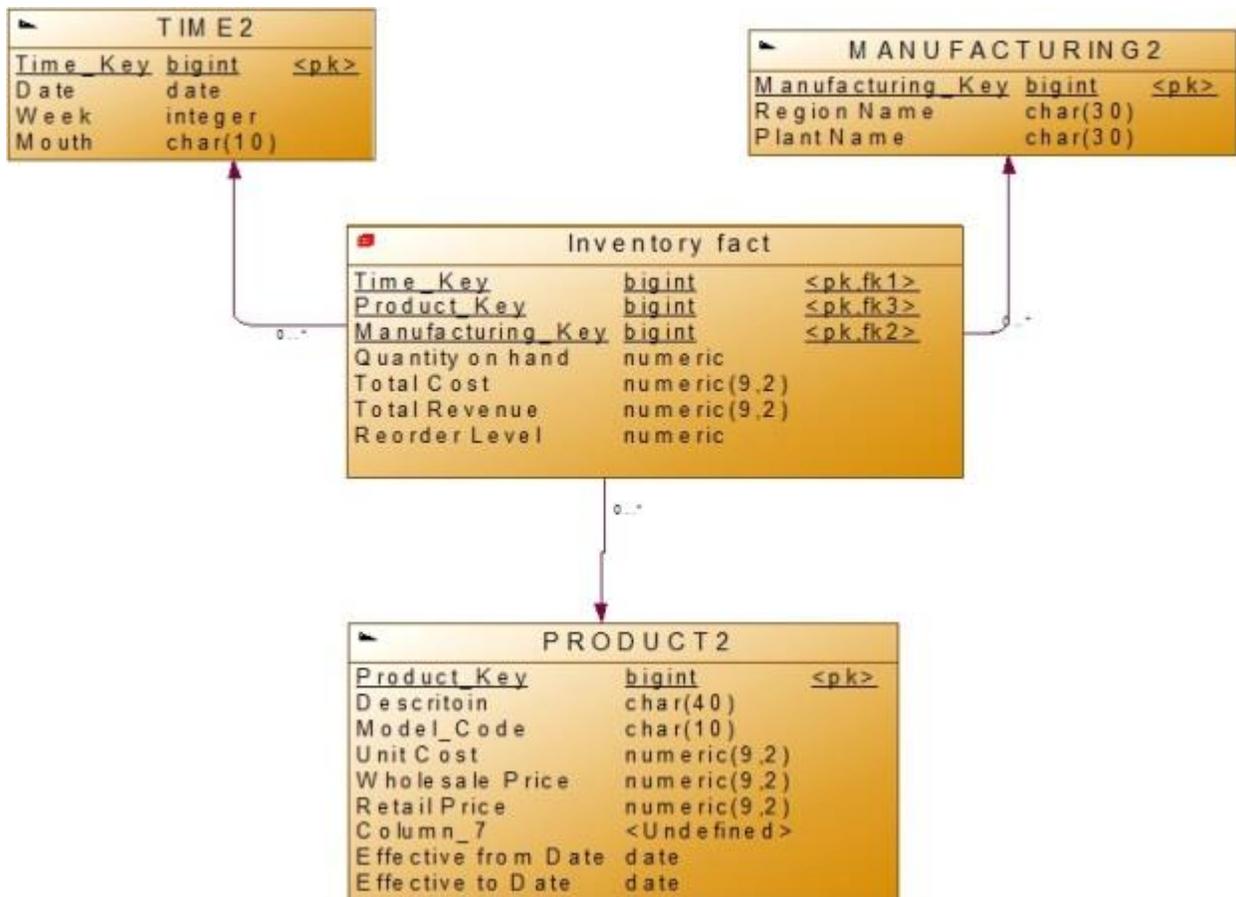
Факты модели ХД

Inventory fact	
Time_Key	bigint
Product_Key	bigint
Manufacturing_Key	bigint
Quantity on hand	numeric
Total Cost	numeric(9,2)
Total Revenue	numeric(9,2)
Reorder Level	numeric

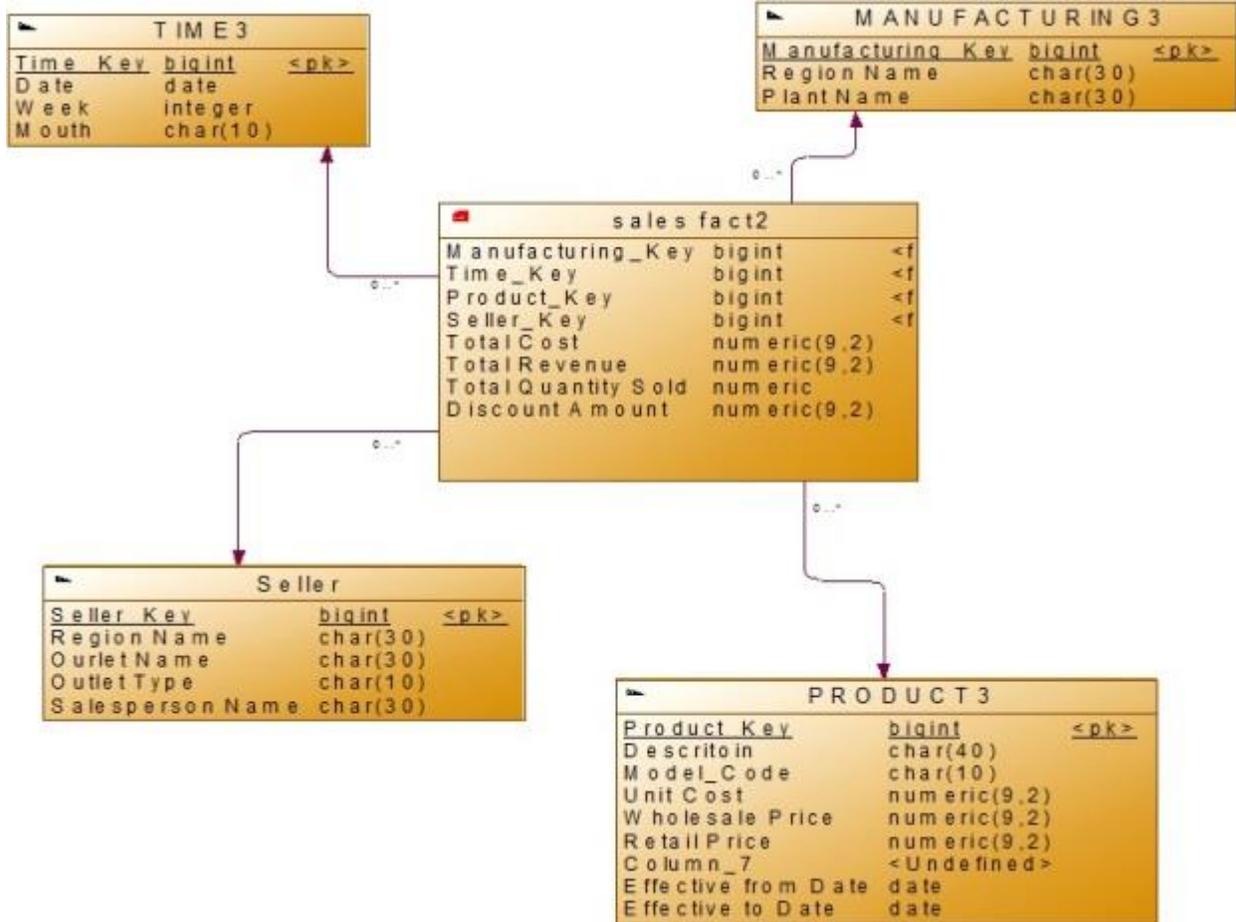
sales fact	
Manufacturing_Key	bigint <1
Sales_Key	bigint
Time_Key	bigint
Product_Key	bigint
Total Cost	numeric(9,2)
Total Revenue	numeric(9,2)
Total Quantity Sold	numeric
Discount Amount	numeric(9,2)

- Из факта 1 получим Inventory fact.
- Из объединенного факта 2 получим sales fact.
- Таким образом, мы имеем два киоска данных.

Первый киоск данных



Второй киоск данных



1.