

Методологии проектирования программных и информационных средств

Лекция 6

Проектирование данных и взаимодействия

Овчинников П.Е.

МГТУ «СТАНКИН»,

ст.преподаватель кафедры ИС

Проектирование баз данных: модели

ER-модель (от англ. *entity-relationship model*, модель «сущность — связь») — модель данных, позволяющая описывать концептуальные схемы предметной области.

ER-модель используется при высокоуровневом (концептуальном) проектировании баз данных. С её помощью можно выделить ключевые сущности и обозначить связи, которые могут устанавливаться между этими сущностями.

Во время проектирования баз данных происходит преобразование ER-модели в конкретную схему базы данных на основе выбранной модели данных (реляционной, объектной, сетевой или др.).

Реляционная модель данных (РМД) — логическая модель данных, прикладная теория построения баз данных, которая является приложением к задачам обработки данных таких разделов математики, как теория множеств и логика первого порядка.

На реляционной модели данных строятся реляционные базы данных.

Проектирование баз данных: уровни моделей

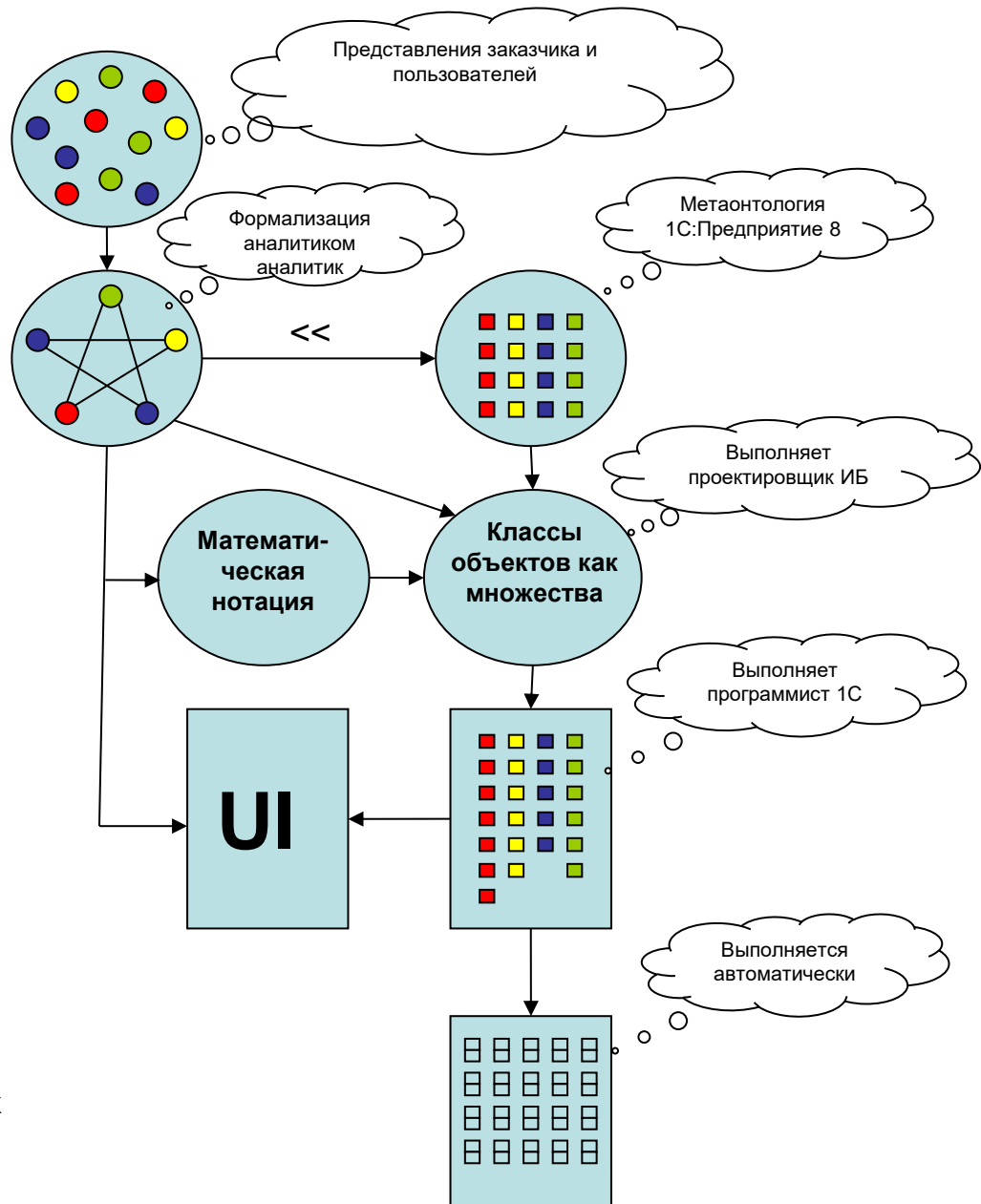
1. Семантическая сеть

2. Онтология

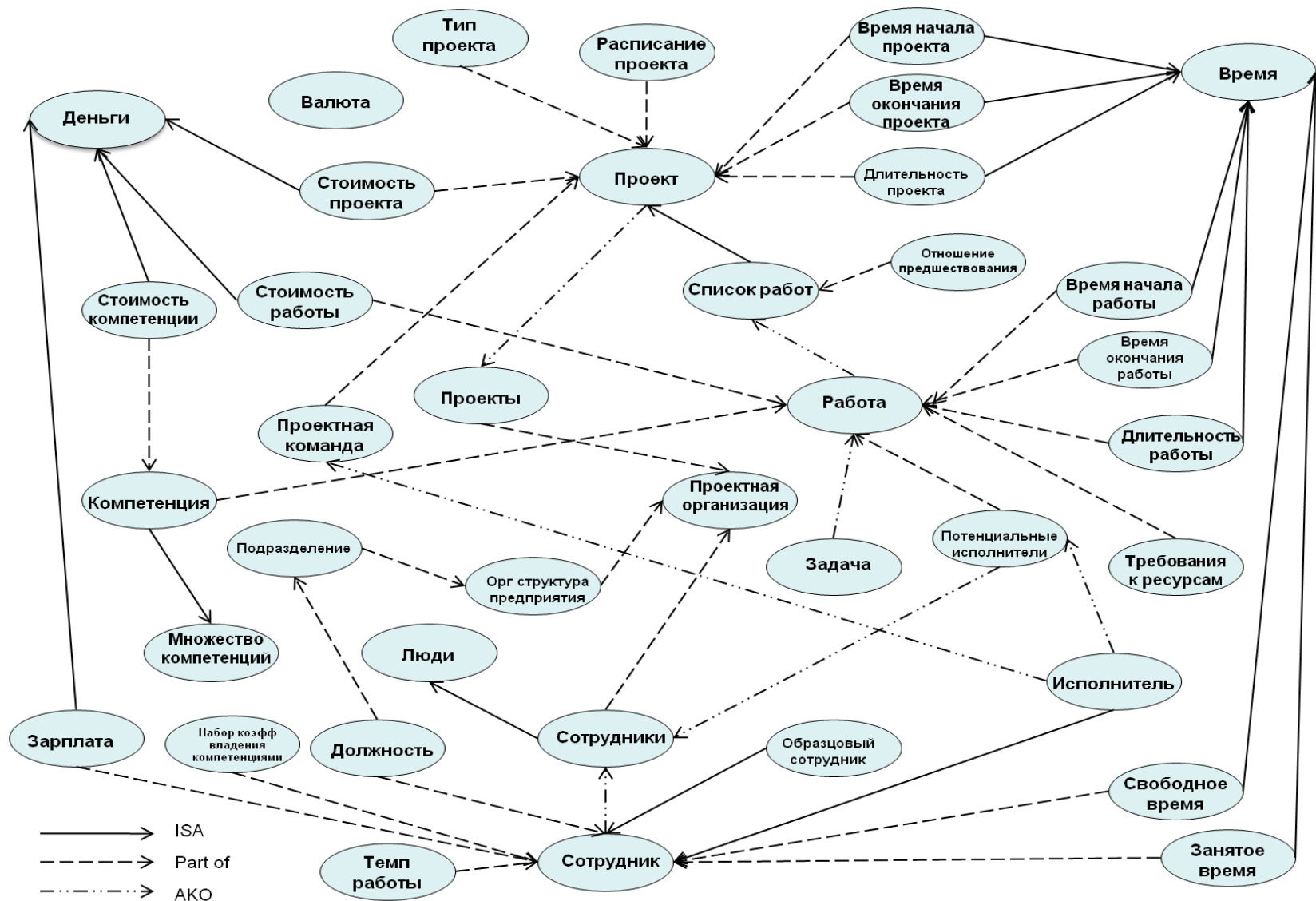
3. Концептуальная модель

4. Объектная модель информационной базы

5. Логическая модель реляционной базы данных



Проектирование баз данных: семантическая сеть



Интероперабельность, полисемия и контекст

ГОСТ Р 55062-2012 Информационные технологии (ИТ). Системы промышленной автоматизации и их интеграция. Интероперабельность. Основные положения

интероперабельность (interoperability): Способность двух или более информационных систем или компонентов к обмену информацией и к использованию информации, полученной в результате обмена

организационная интероперабельность (organizational interoperability): Способность участвующих систем достигать общих целей на уровне бизнес-процессов

семантическая интероперабельность (semantic interoperability): Способность любых взаимодействующих в процессе коммуникации информационных систем одинаковым образом понимать смысл информации, которой они обмениваются

техническая интероперабельность (technical interoperability): Способность к обмену данными между участвующими в обмене системами

барьер интероперабельности (interoperability barrier): Несовместимость сущностей, которая препятствует обмену информацией с другими сущностями, использованию сервисов или общему пониманию обмененных элементов

Интероперабельность, полисемия и контекст

ГОСТ 7.0-99 ИНФОРМАЦИОННО-БИБЛИОТЕЧНАЯ ДЕЯТЕЛЬНОСТЬ, БИБЛИОГРАФИЯ. Термины и определения

синонимия (Synonymy): Свойство двух или нескольких различных по форме, но одинаковых или близких по значению знаков

квазисинонимия (Quasi-synonymy): Свойство двух или нескольких различных знаков, которые могут для особых целей рассматриваться как синонимы

полисемия (Polysemy): Наличие разных, но в какой-либо мере связанных интерпретаций одного и того же знака

омонимия (Homonymy): Свойство двух или более знаков, заключающееся в том, что они имеют одну и ту же материальную форму, но независимые значения

информация (Information): Сведения, воспринимаемые человеком и (или) специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации

коммуникация (Communication): Управляемая передача информации между двумя или более лицами и (или) системами

Интероперабельность, полисемия и контекст

Конте́кст (от лат. *contextus* — "соединение", "связь") — термин, широко используемый в ряде гуманитарных наук (лингвистика, семиотика, социология, философия, антропология), прямо или косвенно изучающих язык и общение. Существует два основных значения термина:

Вербальный контекст —

это **законченный отрывок** письменной или устной речи (текста), **общий смысл** которого позволяет **уточнить значение** входящих в него отдельных слов или предложений. Наличие вербального контекста всегда напрямую влияет на понимание любых сообщений, и отсюда часто встречающаяся в обществе ошибка цитирования, называемая «**вырывание из контекста**» (повторение некоторой усечённой части первоначального текста в ущерб его цельности, что может серьёзно исказить его исходное значение).

Ситуативный контекст —

это **ситуация коммуникации**, включающая **обстановку, время и место**, а также любые факты реальной действительности, помогающие более точно интерпретировать значение высказывания. В этой связи в различных текстах или речах часто упоминается различного рода социальный, культурный, политический или исторический контекст написанного или сказанного. Иногда для описания ситуативного контекста используют лингвистические термины «фоновые знания» или «пресуппозиция».

Полисемия: транзакции (экономика)

Экономика: Теория транзакционных издержек

Теория транзакционных издержек ([англ.](#) transaction cost theory) считается составной частью [Новой институциональной теории](#) и представляет собой теорию организации предприятий, объектом изучения которой служит многосторонний договор как форма организации.

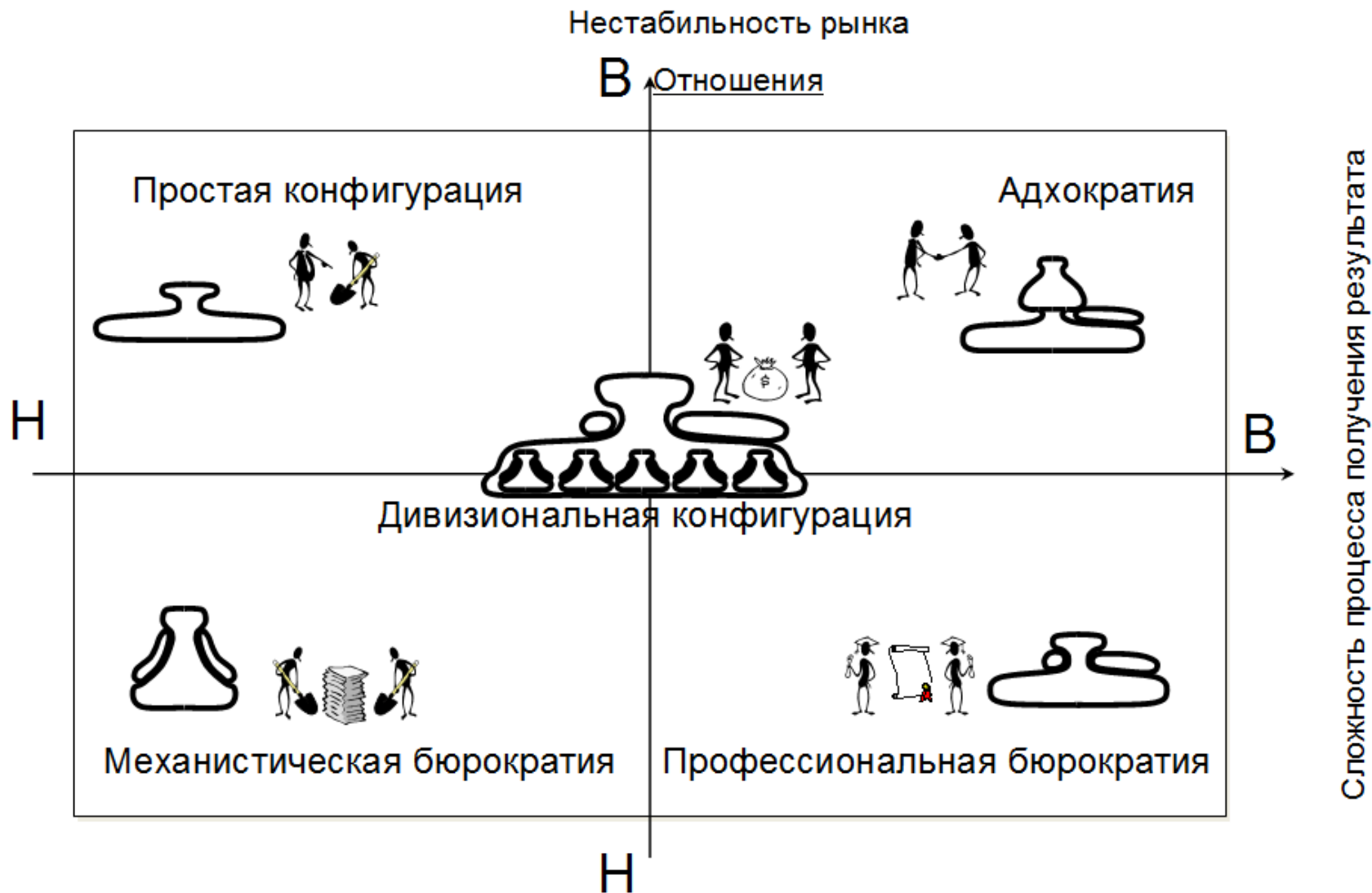
Задачей теории транзакционных издержек является объяснение проблем эффективности тех или иных экономических операций в определённых институциональных рамках, то есть способность различных организационных форм в результативном планировании и осуществлении экономических целей. В основе данной теории находится предположение, что любое действие в экономическом контексте в первую очередь связано с затратами.

Транзакционные издержки — [затраты](#), возникающие в связи с заключением контрактов (в том числе с использованием рыночных механизмов); издержки, сопровождающие взаимоотношения экономических агентов. Выделяют следующие издержки:

- сбора и обработки информации
- проведения переговоров и принятия решений
- контроля
- юридической защиты выполнения

Полисемия: транзакции (экономика)

Стратегическое управление: Теорема Коуза и Конфигурации Минцберга



Полисемия: транзакции (психология)

Трансакция — это единица общения, которая состоит из стимула и реакции.

Трансакционный анализ (*транзакционный анализ, транзактный анализ, транзактный анализ*; сокр. ТА) представляет собой психологическую модель, служащую для описания и анализа поведения человека как индивидуально, так и в составе групп. Данная модель включает философию, теорию и методы, позволяющие людям понять самих себя и особенность своего взаимодействия с окружающими.

Краеугольным камнем транзакционного анализа является положение о том, что один и тот же человек, находясь в некоей определённой ситуации, может функционировать, исходя из одного из трёх эго-состояний, чётко отличимых одно от другого:

Эго-состояние Родителя (Р) содержит установки и поведение, перенятые извне, в первую очередь — от родителей. Внешне они часто выражаются в предубеждениях, критическом и заботливом поведении по отношению к другим. Внутренне — переживаются как старые родительские наказы, которые продолжают влиять на нашего внутреннего Ребёнка.

Эго-состояние Взрослого (В) не зависит от возраста личности. Оно ориентировано на восприятие текущей реальности и на получение объективной информации. Оно является организованным, хорошо приспособленным, находчивым и действует, изучая реальность, оценивая свои возможности и спокойно рассчитывая.

Эго-состояние Ребёнка (Ре) содержит все побуждения, которые возникают у ребёнка естественным образом. Оно также содержит запись ранних детских переживаний, реакций и позиций в отношении себя и других. Оно выражается как «старое» (архаическое) поведение детства. Эго-состояние Ребёнка отвечает также за творческие проявления личности.

Полисемия: транзакции (информатика)

Транзакция (англ. *transaction*) — группа последовательных операций с базой данных, которая представляет собой логическую единицу работы с данными.

Транзакция может быть выполнена либо целиком и успешно, соблюдая целостность данных и независимо от параллельно идущих других транзакций, либо не выполнена вообще, и тогда она не должна произвести никакого эффекта.

Транзакции обрабатываются транзакционными системами, в процессе работы которых создаётся история транзакций.

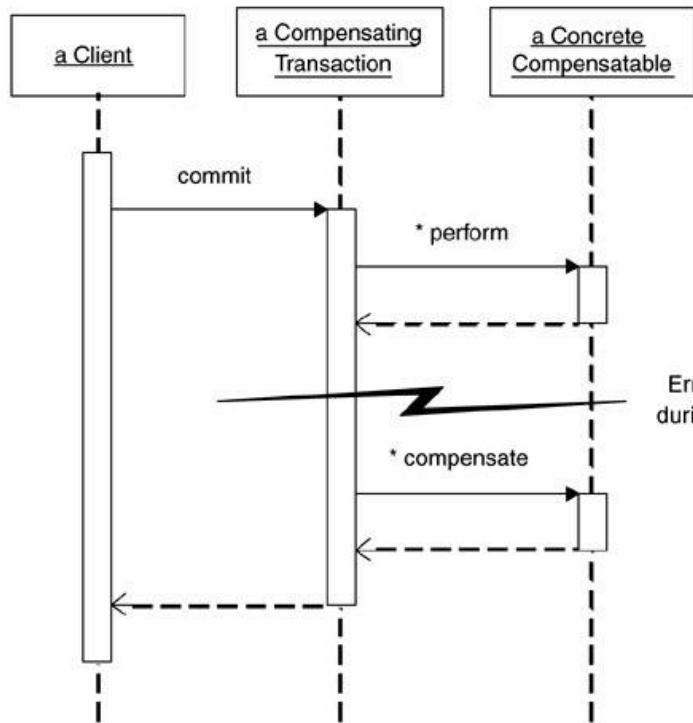
Различают последовательные (обычные), параллельные и распределённые транзакции.

Распределённые транзакции подразумевают использование более чем одной транзакционной системы и требуют намного более сложной логики (например, two-phase commit — двухфазный протокол фиксации транзакции).

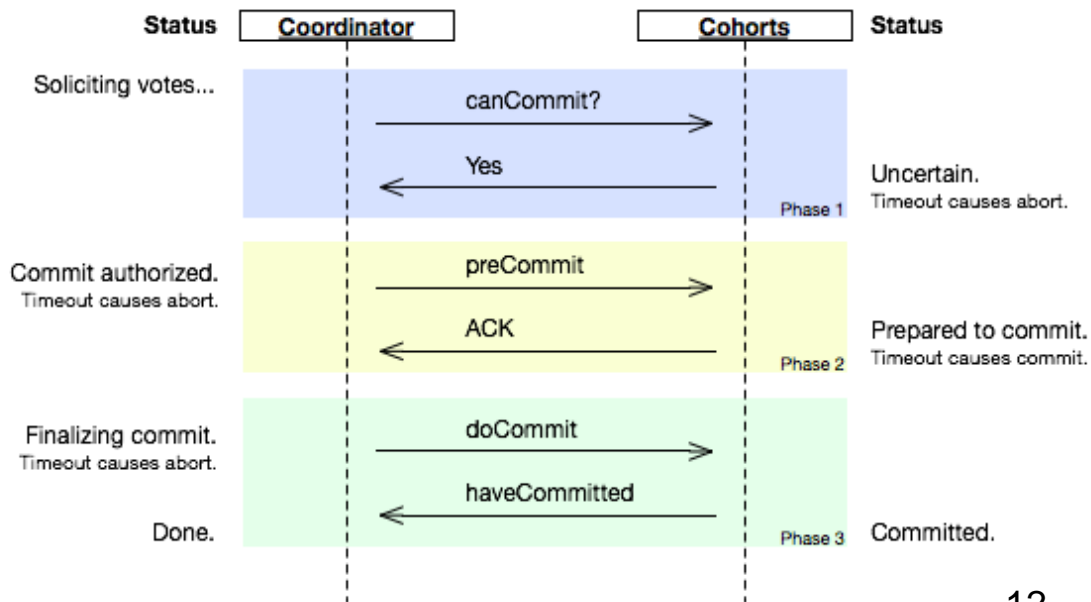
Также в некоторых системах реализованы автономные транзакции, или подтранзакции, которые являются автономной частью родительской транзакции.

Транзакции

Отмена (abort) и компенсация (compensation)



Двухфазная обработка



Трехфазная обработка

Проектирование баз данных: технологии

OLAP ([англ. online analytical processing](#), интерактивная аналитическая обработка) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу. Реализации технологии OLAP являются компонентами программных решений класса [Business Intelligence^{\[1\]}](#).

Существуют три типа OLAP:^[2]

многомерная OLAP (Multidimensional OLAP — [MOLAP](#));

реляционная OLAP (Relational OLAP — [ROLAP](#));

гибридная OLAP (Hybrid OLAP — [HOLAP](#)).

NoSQL ([англ. not only SQL, не только SQL](#)), в [информатике](#) — термин, обозначающий ряд подходов, направленных на реализацию хранилищ [баз данных](#), имеющих существенные отличия от моделей, используемых в традиционных [реляционных СУБД](#) с доступом к данным средствами языка [SQL](#).

Применяется к базам данных, в которых делается попытка решить проблемы [масштабируемости](#) ([англ. scalability](#)) и [доступности](#) ([англ. availability](#)) за счёт [атомарности](#) ([англ. atomicity](#)) и [согласованности данных](#) ([англ. consistency](#))^[1].

Проектирование баз данных: технологии

Объектно-реляционная СУБД (ОРСУБД) — [реляционная СУБД \(РСУБД\)](#), поддерживающая некоторые технологии, реализующие [объектно-ориентированный подход](#): объекты, классы и наследование реализованы в структуре баз данных и языке запросов.

Объектно-реляционными [СУБД](#) являются, например, широко известные [Oracle Database](#), [Informix](#), [DB2](#), [PostgreSQL](#).

Object-Oriented Model

Object 1: Maintenance Report **Object 1 Instance**

Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

01-12-01
24
I-95
2.5
6.0
6.0

Object 2: Maintenance Activity

Activity Code
Activity Name
Production Unit
Average Daily Production Rate

Example of XML Type Query in IBM DB2 SQL [\[edit\]](#)

```
select
  id, vol, xmlquery('$j/name', passing journal as "j") as name
from
  journals
where
  xmlexists('$j[licence="CreativeCommons"]', passing journal as "j")
```

OLTP-системы

OLTP ([англ. Online Transaction Processing](#)), **транзакционная система** — обработка транзакций в реальном времени.

Способ организации базы данных, при котором система работает с небольшими по размерам транзакциями, но идущими большим потоком, и при этом клиенту требуется от системы минимальное время отклика.

Термин OLTP применяют также к информационным системам и приложениям. OLTP-системы предназначены для ввода, структурированного хранения и обработки информации (операций, документов) в режиме реального времени.

OLTP-приложениями охватывается широкий спектр задач во многих отраслях — [автоматизированные банковские системы](#), [ERP-системы \(системы планирования ресурсов предприятия\)](#), банковские и биржевые операции, в промышленности — регистрация прохождения детали на [конвейере](#), фиксация в статистике посещений очередного посетителя веб-сайта, автоматизация бухгалтерского, складского учёта и учёта документов и т. п. Приложения OLTP, как правило, автоматизируют структурированные, повторяющиеся задачи обработки данных, такие как ввод заказов и банковские транзакции. OLTP-системы проектируются, настраиваются и оптимизируются для выполнения максимального количества транзакций за короткие промежутки времени. Как правило, большой гибкости здесь не требуется, и чаще всего используется фиксированный набор надёжных и безопасных методов ввода, модификации, удаления данных и выпуска оперативной отчётности. Показателем эффективности является количество транзакций, выполняемых за секунду. Обычно аналитические возможности OLTP-систем сильно ограничены (либо вообще отсутствуют).

OLAP-системы

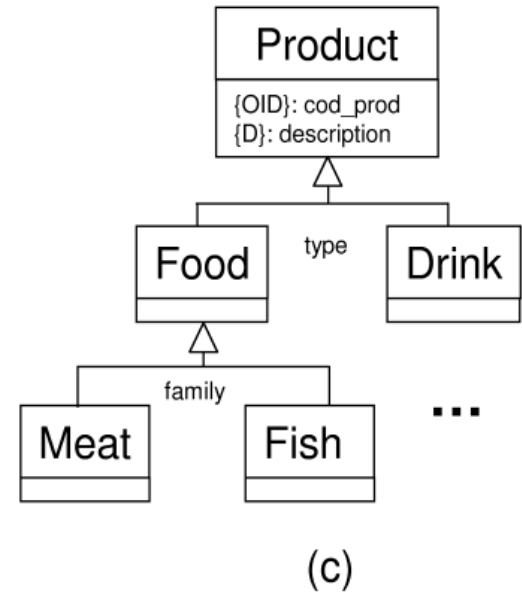
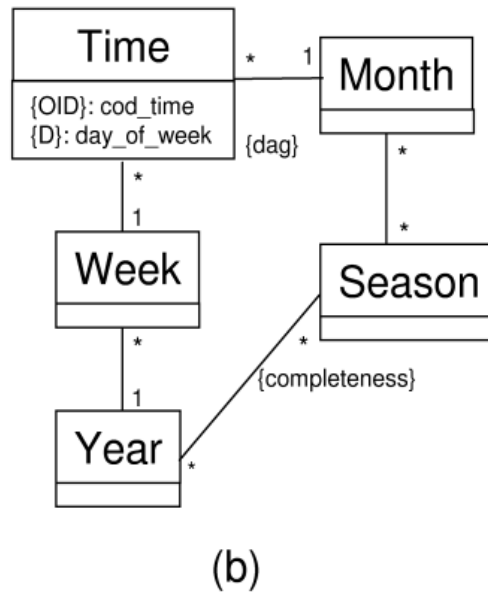
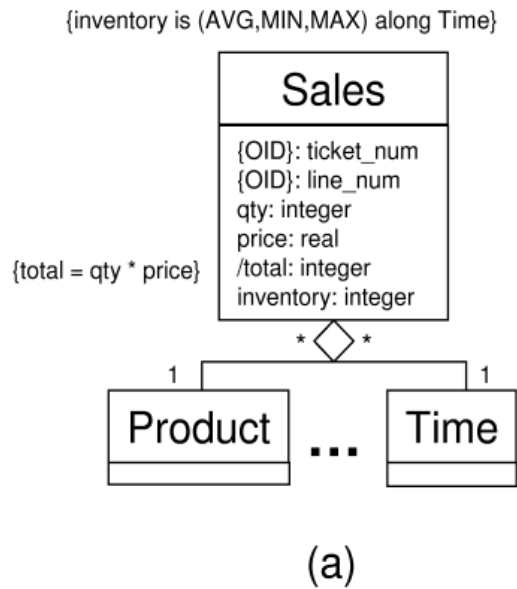
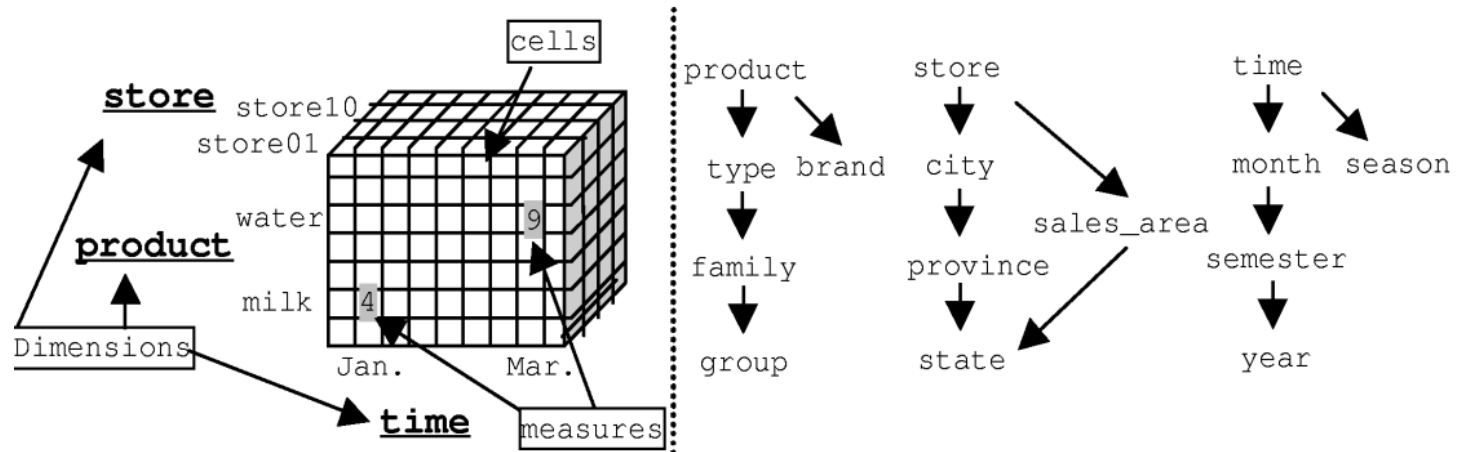
OLAP ([англ. online analytical processing](#), аналитическая обработка в реальном времени) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Реализации технологии OLAP являются компонентами программных решений класса [Business Intelligence](#).

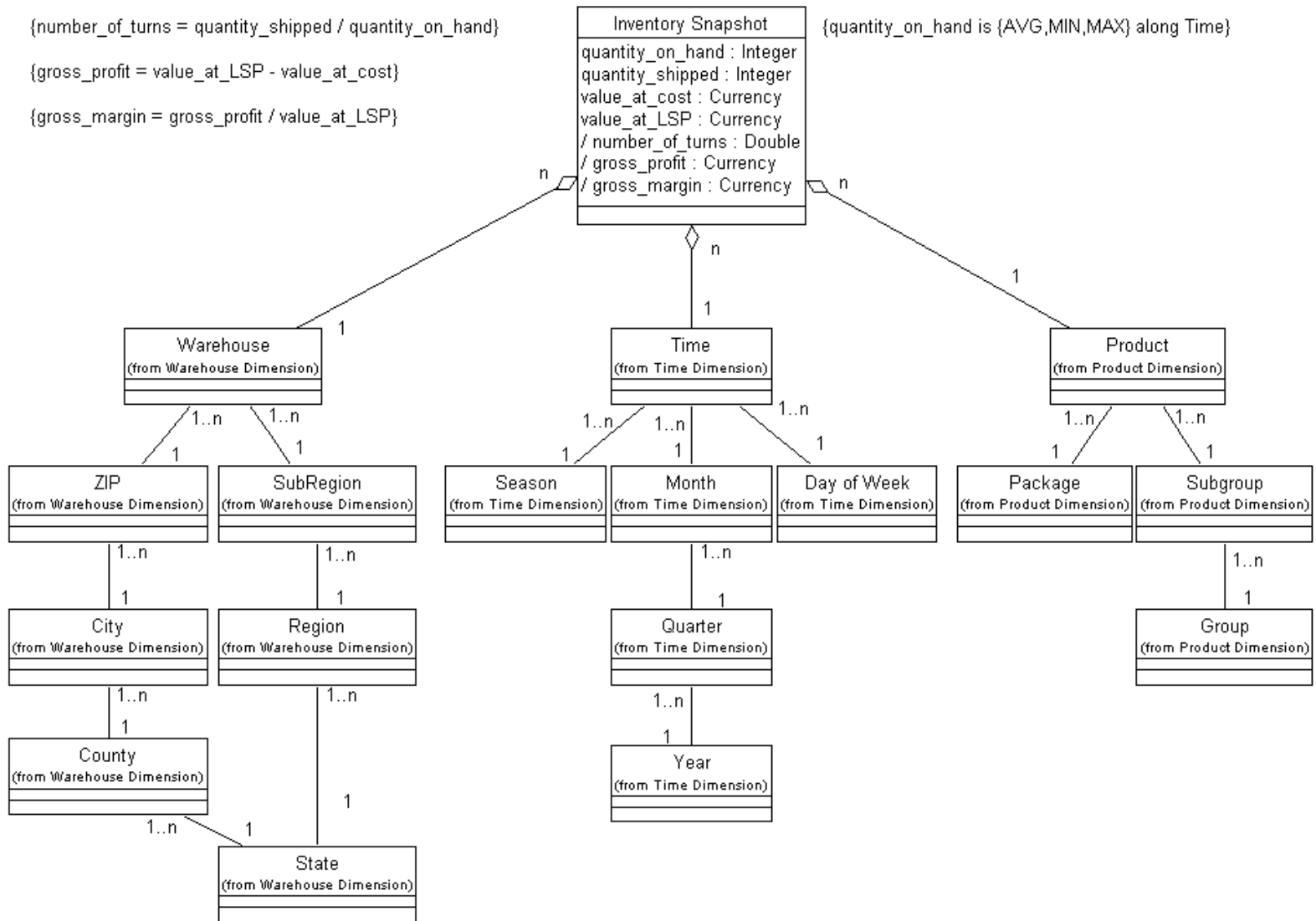
Многомерное моделирование является методом моделирования и визуализации данных как множества числовых или лингвистических показателей или параметров (measures), которые описывают общие аспекты деятельности организации и базируется на следующих основных понятиях:

- **Факт (Fact)**
набор связанных элементов данных, содержащих метрики и описательные данные.
- **Атрибут (Attribute)**
описание характеристики реального объекта предметной области.
- **Измерение (Dimension)**
интерпретация факта с некоторой точки зрения в реальном мире.
- **Параметр, метрика или показатель (Measure)**
числовая характеристика факта
- **Гранулированность (Granularity)**
уровень детализации данных

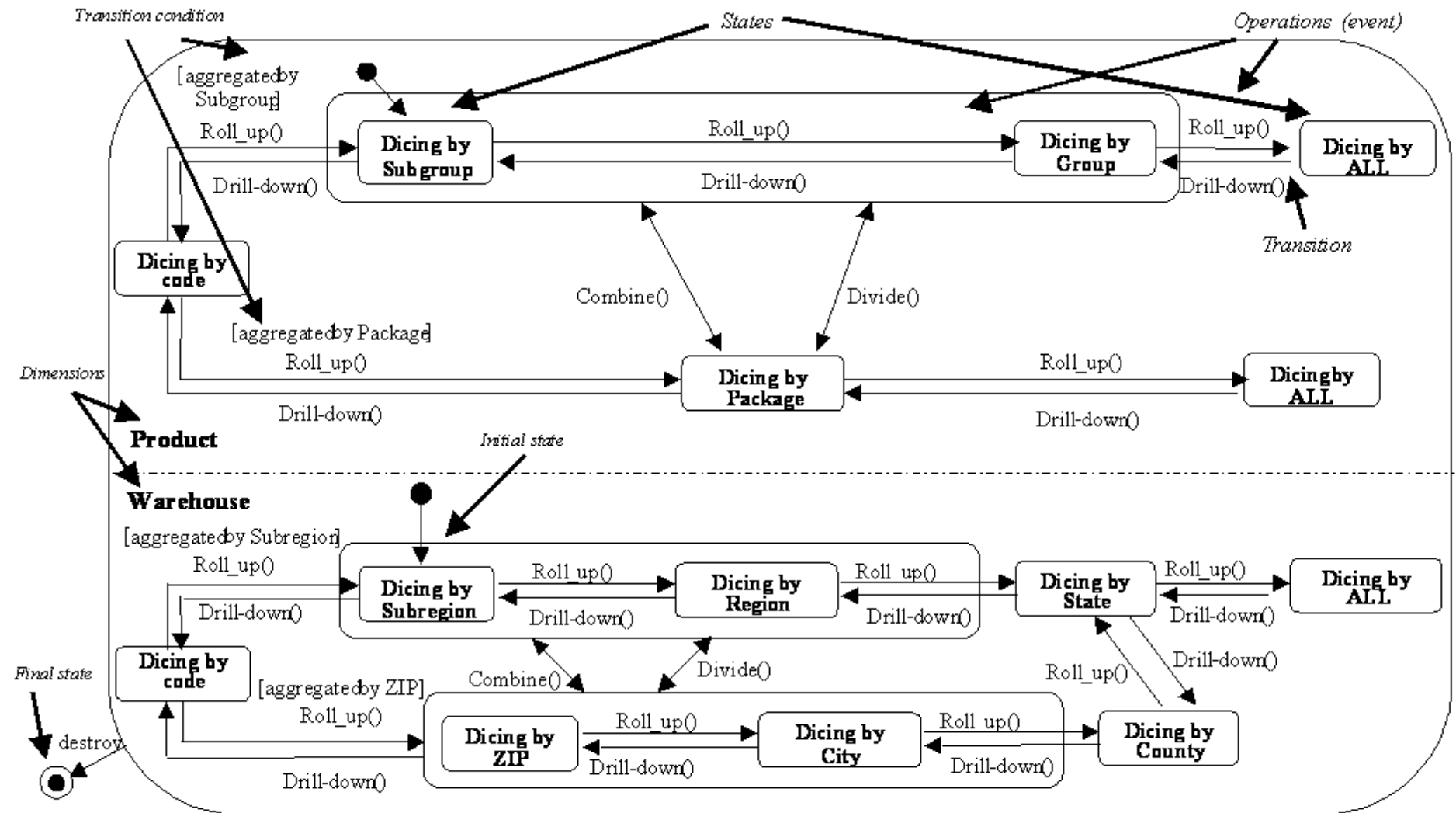
OLAP-системы



OLAP-системы



OLAP-системы



Хранилища данных: Data Vault

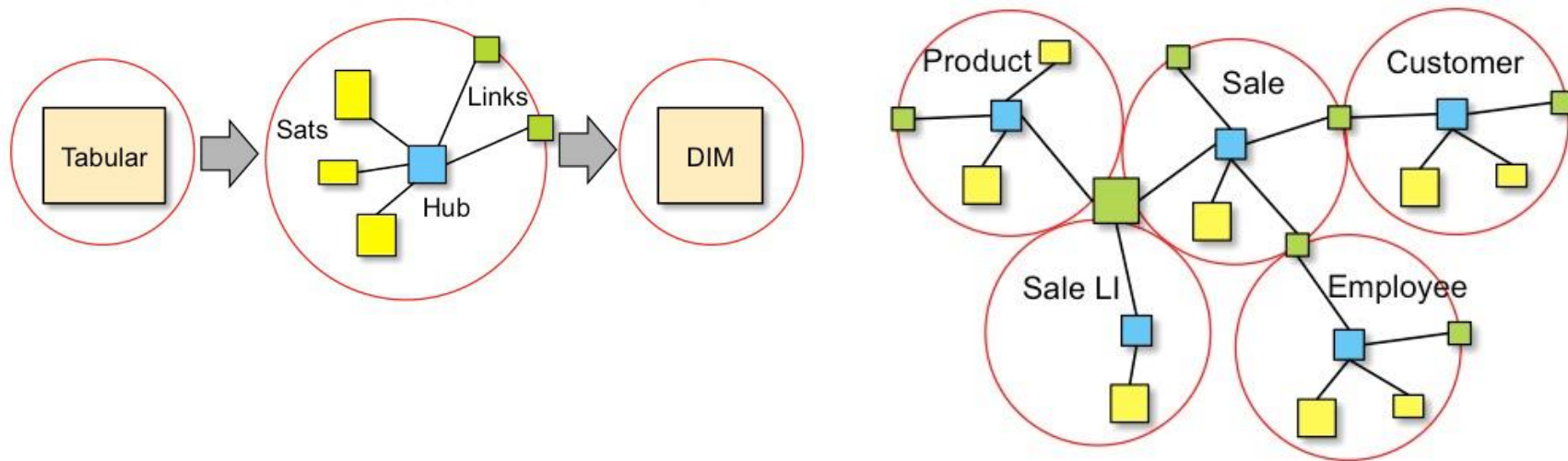
Модель Data Vault представляет собой набор связанных между собой нормализованных таблиц, ориентированных на хранение детализированной информации с возможностью отслеживания происхождения данных и поддерживающих одну или несколько областей бизнеса.

В модели Data Vault используется всего три типа таблиц:

Hub обеспечивает представление функциональных областей предметной области

Link обеспечивает транзакционную связь между Hub-таблицами

Satellite предоставляет детализацию первичного ключа Hub-таблицы



Резидентные базы данных

Резидентная база данных ([англ. in-memory database, IMDB](#)) — [база данных](#), размещаемая в [оперативной памяти](#). **Резидентная СУБД** — [система управления](#) резидентными базами данных, один из видов программных систем, работающих в парадигме [резидентных вычислений](#) ([англ. in-memory computing](#)).

Резидентные СУБД за счёт оптимизаций, возможных в условиях хранения и обработки в байтоадресуемой оперативной памяти, обеспечивают лучшее быстродействие, чем СУБД, работающие с базами данных на устройствах постоянного хранения, как правило, с [блочной](#) организацией, и подключаемых по [шинным](#) или [сетевым](#) интерфейсам. При этом размер резидентной базы данных ограничен ёмкостью оперативной памяти узла. Для ряда резидентных СУБД реализуются техники [репликации](#) и [сегментирования](#) ([англ. sharding](#)), позволяющие работать с единой резидентной базой данных на нескольких узлах. Поскольку оперативная память [энергозависима](#), то используется [запись с предварительным журналированием](#) на энергонезависимом устройстве для обеспечения [целостности базы данных](#) при внезапной [перезагрузке](#), то есть, работа с резидентной базой не исключает зависимости от производительности подсистемы [ввода-вывода](#) (хотя и снижает её).

Системы классификации

Таксономия (от др.-греч. τάξις — строй, порядок и νόμος — закон) — учение о принципах и практике классификации и систематизации

Математически таксономией является древообразная структура классификаций определенного набора объектов. Вверху этой структуры — объединяющая единая классификация — корневой таксон — которая относится ко всем объектам данной таксономии. Таксоны, находящиеся ниже корневого, являются более специфическими классификациями, которые относятся к поднаборам общего набора классифицируемых объектов

Система классификации Линнея

Все живые существа можно классифицировать с помощью иерархической системы, в основе которой лежат категории рода и вида.

В системе Линнея все относится:

- либо к животным,
- либо к растениям,
- либо к неживой природе (минералам).

Таксономия системы:

Царство -> Тип -> Класс -> Отряд -> Семейство -> Род -> Вид



Системы кодирования

Система кодирования - строго определенный порядок присвоения условных обозначений единицам информации. Таким образом, все коды строятся по определенным правилам (системам).

Используемые для этих целей системы построения кодов подразделяются на:

- линейные (одномерные);
- шахматные.

Линейные системы кодирования

В линейных кодах условное обозначение соответствует только одной единице информации. По способу построения различают следующие линейные системы кодирования:

- порядковые,
- серийные,
- позиционные (разрядные, или десятичные),
- повторения,
- смешанные (комбинированные).

Шахматная система кодирования

В шахматных используются двухпозиционные коды, одновременно отражается характеристика двух информационных единиц (по строке и столбцу).

Системы кодирования

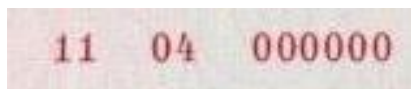
Порядковая система кодирования предполагает последовательное присвоение условных обозначений кодируемым единицам информации.

Специальной классификации информации, как правило, не требуется.

Последовательность кодов задается, прежде всего, хронологией возникновения информационных единиц, но чаще всего объектом кодирования выступает информация, упорядоченная (систематизированная) по алфавиту.

Серийная система кодирования ориентирована на разделение классифицируемого множества по какому-либо признаку на отдельные части (серии). За каждой серией закрепляется своя группа условных обозначений (чисел, называемых номерами).

При этом номера единиц информации последующих серий не продолжают последовательно номера уже имеющихся единиц предыдущей серии, в результате создается определенный разрыв номеров, используемый в качестве резерва для последующего расширения (в случае необходимости) множества кодируемых позиций в каждой серии без нарушения общей логики построения списка.

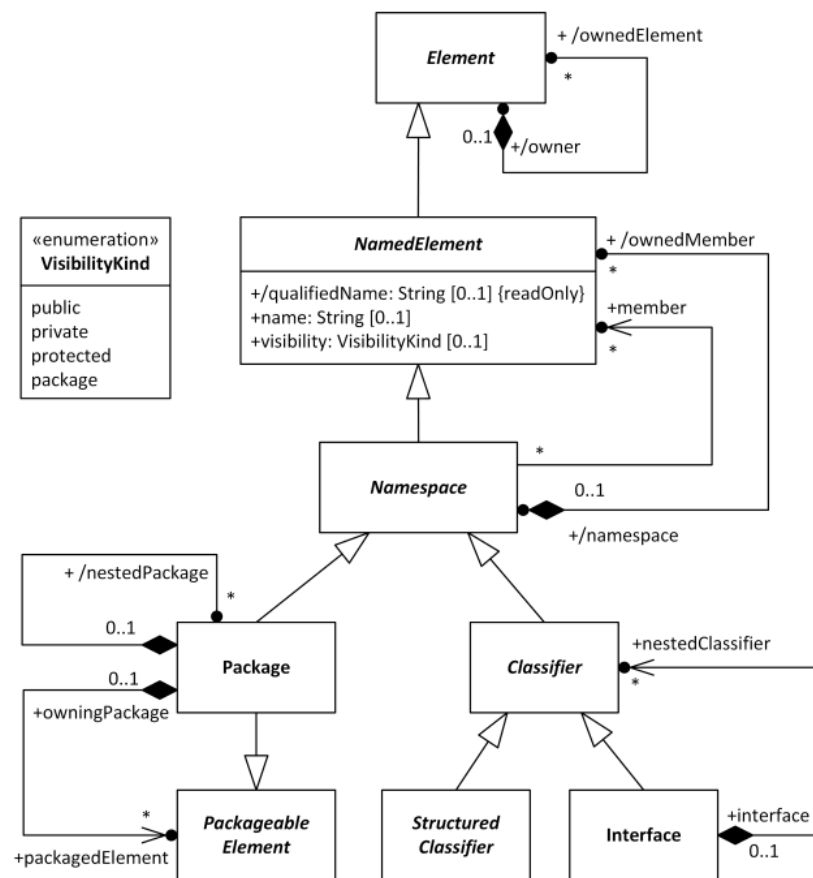


Системы кодирования

Позиционная (разрядная, десятичная) система кодирования предполагает иерархическую структуру представления информации или разделение ее по нескольким соподчиненным признакам.

Сущность данной системы заключается в том, что каждый уровень (или признак) классификации обеспечивается своей нумерацией в пределах всего уровня или признака (группы информации). При этом устанавливается предел разрядности группы и выбирается ее некоторая кратность.

Позиционная система применяется для кодирования сложных составных (иерархических) номенклатур, в которых, как правило, каждый реквизит, характеризующий низший уровень классификации, получает ряд характеристик, отражающих его принадлежность к более высокому уровню классификации.



Переходные ключи

Гармонизация общероссийского классификатора: Приведение общероссийского классификатора в соответствие с международной (региональной) классификацией, межгосударственным классификатором или международным (региональным) стандартом по классификации установленными путями гармонизации.

Переходной ключ: Таблица, устанавливающая соответствие каждой группировки или объекта классификации общероссийского классификатора одной или нескольким группировкам или объектам сопоставляемой классификации.

Код		Дробная единица ^[с 1]	Наименования валют ^[с 2]		Наименования государств и территорий ^[с 3]
буквенный	цифровой	разряды	ОКВ	ISO 4217	в соответствии с ISO 4217 и/или ОКВ
Денежные единицы, включённые в действующую редакцию ОКВ					
AED	784	2	Дирхам (ОАЭ)	UAE Dirham	 ОАЭ
AFN	971	2	Афгани	Afghani	 Афганистан
ALL	008	2	Лек	Lek	 Албания
AMD ^[a 1]	051	2	Армянский драм	Armenian Dram	 Армения
ANG	532	2	Нидерландский антильский гульден	Netherlands Antillean Guilder	 Кюрасао  Синт-Мартен
AOA	973	2	Кванза	Kwanza	 Ангола
ARS	032	2	Аргентинское песо	Argentine Peso	 Аргентина

НСИ, MDM

ГОСТ 34.003-90 Информационные технологии. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения

Нормативно-справочная информация (НСИ) автоматизированной системы - Информация, заимствованная из нормативных документов и справочников и используемая при функционировании

Управление основными данными, управление мастер-данными (англ. *Master Data Management*, MDM) — совокупность процессов и инструментов для постоянного определения и управления основными данными компании (в том числе справочными).

Мастер-данные — это данные с важнейшей для ведения бизнеса информацией: о клиентах, продуктах, услугах, персонале, технологиях, материалах и так далее. Они относительно редко изменяются и не являются транзакционными.

Взаимодействие (философия)

Взаимодей́ствие — базовая философская категория, отражающая **процессы** воздействия **объектов** (**субъектов**) друг на друга, их изменения, взаимную обусловленность и порождение одним объектом других

По сути, взаимодействие представляет собой разновидность **опосредованной** или **непосредственной**, **внутренней** или **внешней связи**; при этом **свойства** любых **объектов** могут быть познанными или проявить себя только во взаимодействии с другими **объектами**

Философское понятие *взаимодействия*, нередко выступая в роли интеграционного фактора, обуславливает **объединение** отдельных **элементов в некий новый вид** целостности, и, таким образом, имеет глубокую связь с понятием структуры

Взаимодействие — объективная и универсальная форма движения, развития, которая определяет существование и **структурную организацию** любой материальной системы

Воздействие (управление)

Система управления — систематизированный (строго определённый) набор средств:

- **сбора сведений** о подконтрольном **объекте** и
 - средств **воздействия** на его поведение
- предназначенный для достижения определённых целей

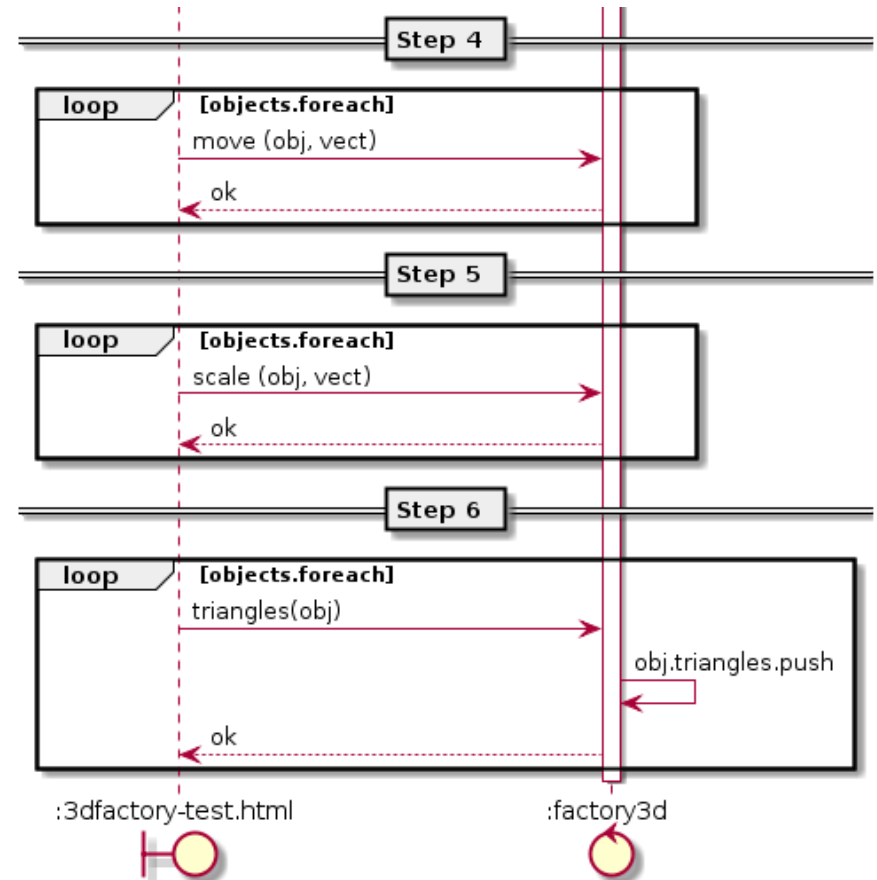
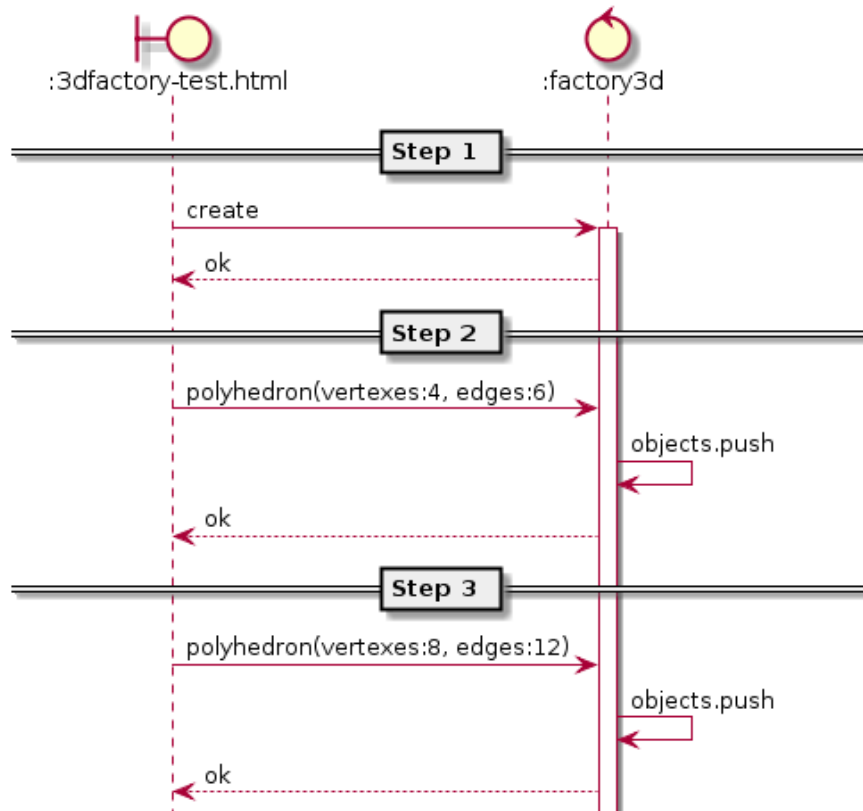
Объектом системы управления могут быть как технические объекты, так и люди. Объект системы управления может состоять из других объектов, которые могут иметь постоянную структуру взаимосвязей.

Техническая структура управления — устройство или набор устройств для манипулирования поведением других устройств или систем

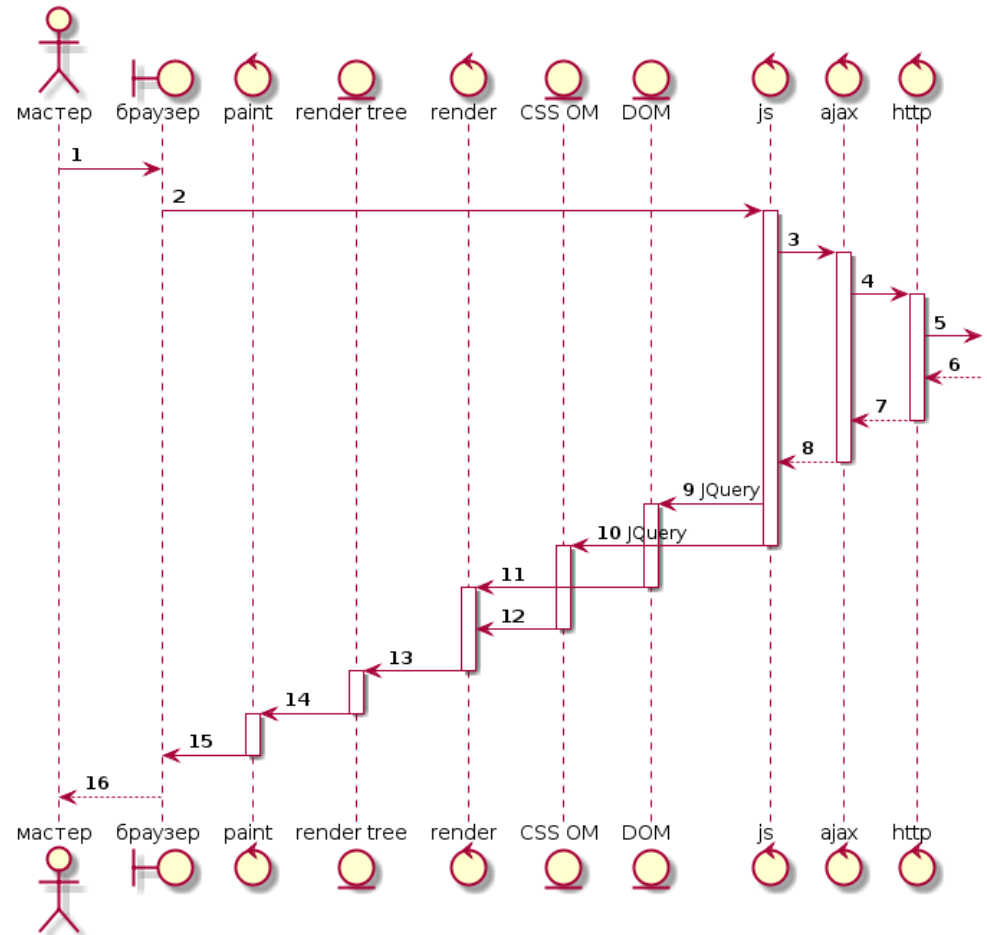
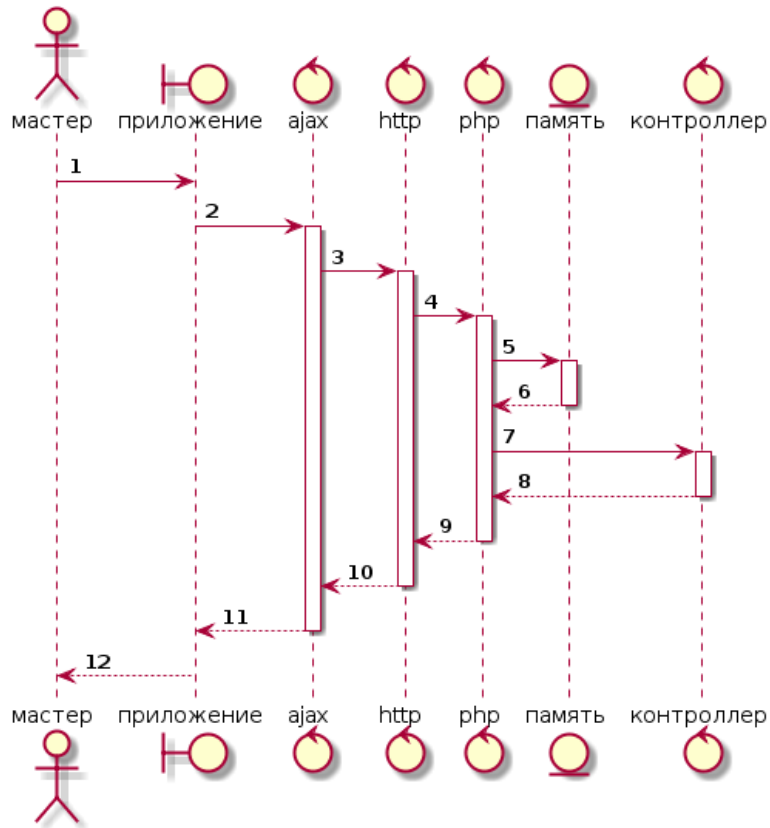
Объектом управления может быть любая динамическая система или её модель

Состояние объекта характеризуется некоторыми количественными величинами, изменяющимися во времени, то есть переменными состояниями

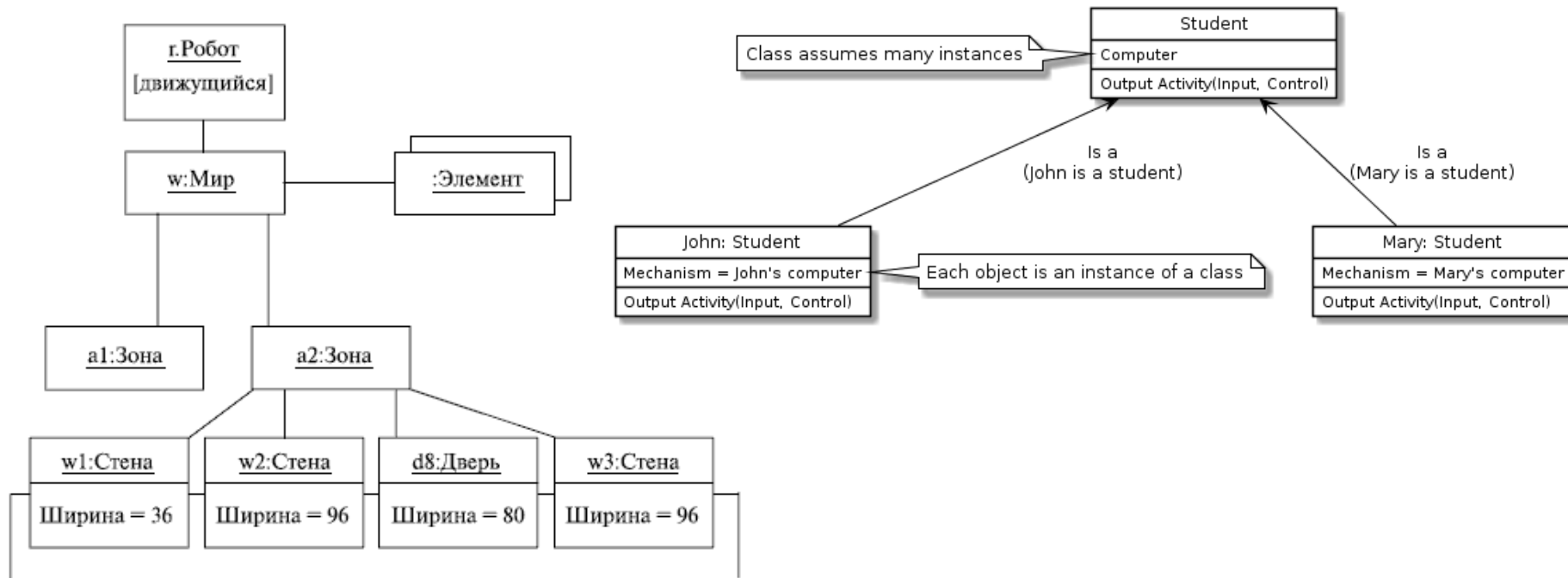
Проектирование взаимодействия: диаграммы



Проектирование взаимодействия: актеры



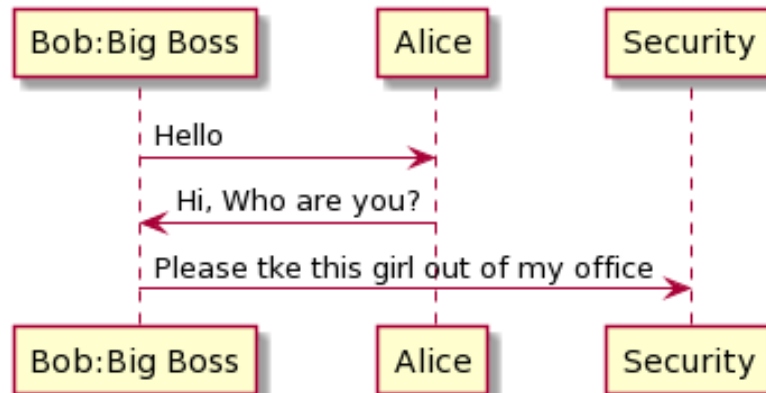
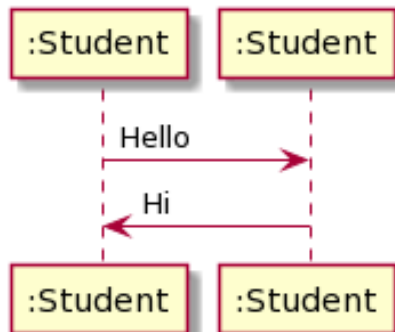
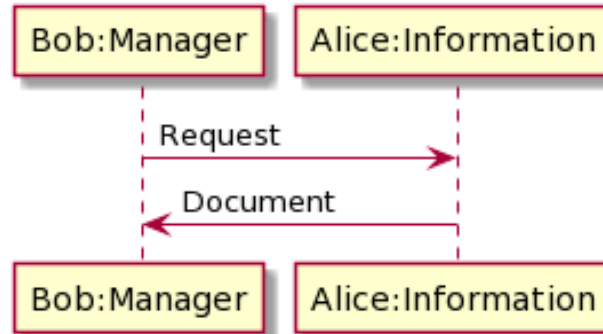
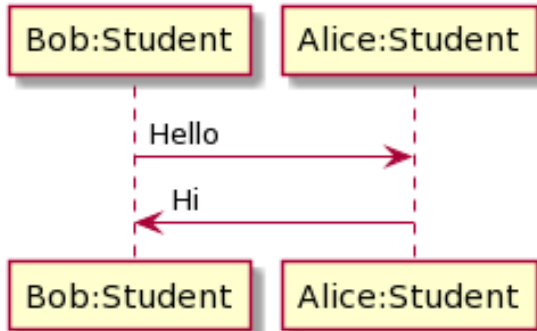
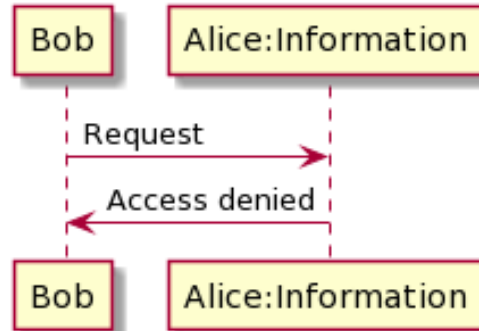
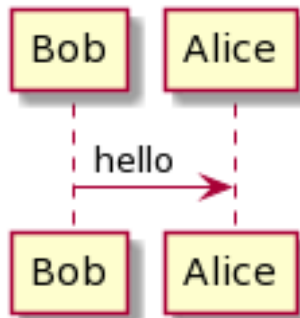
Проектирование взаимодействия: объекты



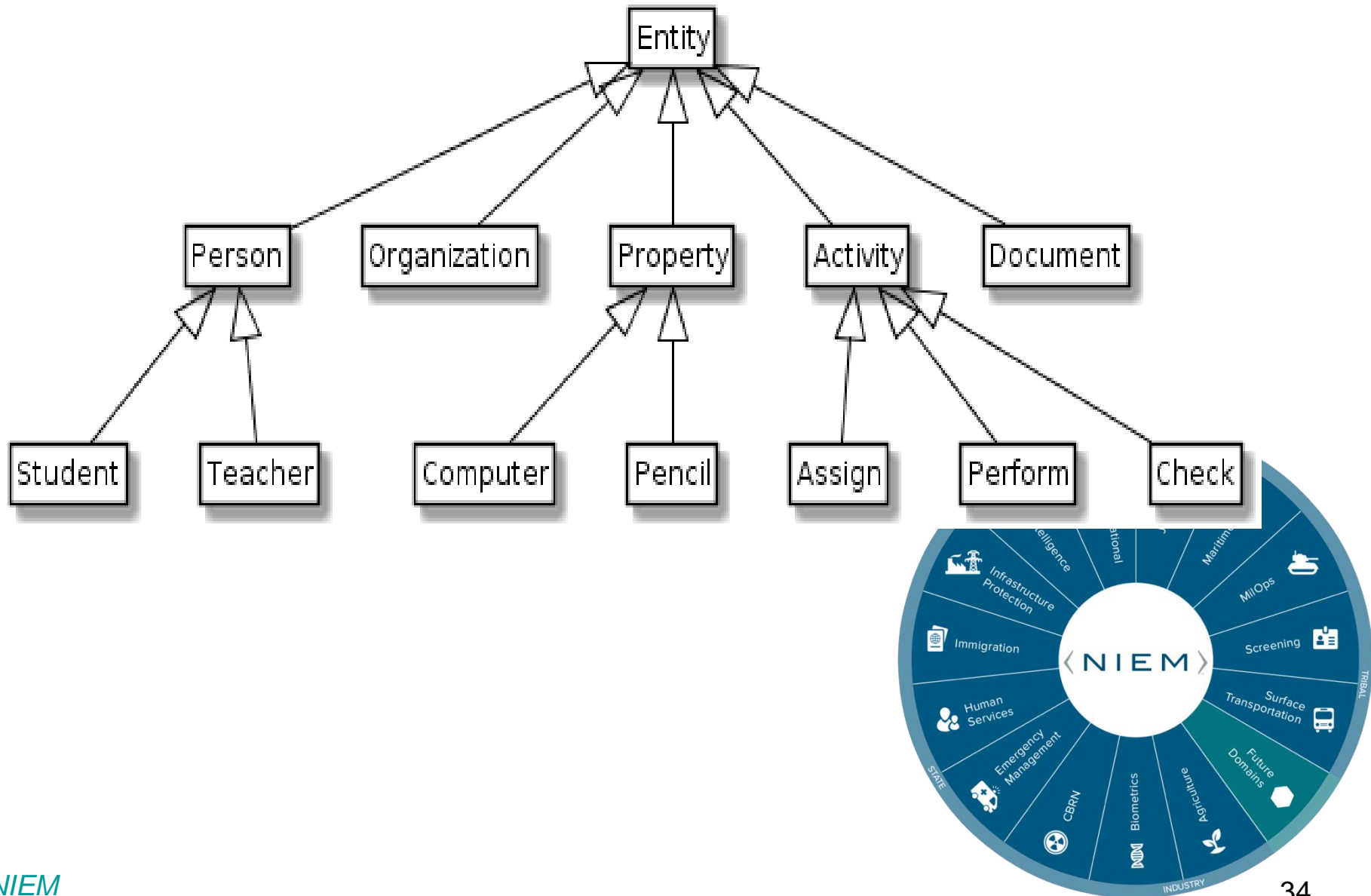
Объекты в UML принято обозначать прямоугольником с двумя секциями. В верхней секции помещается имя объекта и название его класса, и первое и второе выделяется подчёркиванием. **Имя объекта может быть опущено.**

Формат названия в общем виде имеет следующий вид:
имя объекта : имя класса

Проектирование взаимодействия: субъекты



Проектирование взаимодействия: онтологии



Проектирование взаимодействия

The Global JXDM endeavor began in March 2001 as a reconciliation of data definitions and evolved into a broad two-year effort to develop an XML-based framework that would enable the entire justice and public safety communities to effectively share information at all levels--laying the foundation for local, state, tribal, and national justice interoperability.

Approximately **16,000 justice and public safety-related data elements** were collected from various local and state government sources. These were analyzed and **reduced to around 2,000 unique data elements** that were then incorporated into about **300 data objects**, or reusable components, resulting in the Global Justice XML Data Dictionary (Global JXDD).

The Global JXDD components have inherent qualities enabling access from multiple sources and reuse in multiple applications. In addition, the standardization of the core components resulted in significant potential for increased interoperability among and between justice and public safety information systems.

ООП: определение данных

XML Schema

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="country">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="country_name" type="xs:string"/>
        <xs:element name="population" type="xs:decimal"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```
<grammar xmlns="http://relaxng.org/ns/structure/1.0">
  <start>
    <element name="book">
      <oneOrMore>
        <ref name="page"/>
      </oneOrMore>
    </element>
  </start>
  <define name="page">
    <element name="page">
      <text/>
    </element>
  </define>
</grammar>
```

ООП: определение данных

XML

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE recipe>
<recipe name="хлеб" preptime="5min" cooktime="180min">
  <title>
    Простой хлеб
  </title>
  <composition>
    <ingredient amount="3" unit="стакан">Мука</ingredient>
    <ingredient amount="0.25" unit="грамм">Дрожжи</ingredient>
    <ingredient amount="1.5" unit="стакан">Тёплая вода</ingredient>
  </composition>
  <instructions>
    <step>
      Смешать все ингредиенты и тщательно замесить.
    </step>
    <step>
      Закрыть тканью и оставить на один час в тёплом помещении.
    </step>
  </instructions>
</recipe>
```

```
<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main">
  <w:body>
    <w:p>
      <w:r>
        <w:t>Example text.</w:t>
      </w:r>
    </w:p>
  </w:body>
</w:document>
```

ООП: определение данных

JSON

```
{
  "firstName": "Иван",
  "lastName": "Иванов",
  "address": {
    "streetAddress": "Московское ш., 101, кв.101",
    "city": "Ленинград",
    "postalCode": 101101
  },
  "phoneNumbers": [
    "812 123-1234",
    "916 123-4567"
  ]
}
```

```
{
  "array": [
    1,
    2,
    3
  ],
  "boolean": true,
  "color": "gold",
  "null": null,
  "number": 123,
  "object": {
    "a": "b",
    "c": "d"
  },
  "string": "Hello World"
}
```