CS5200
Database Management

Ken Baclawski
Spring 2017

Welcome

- Introductions
- Professor

  Ken Baclawski <k.baclawski@neu.edu>
- Teaching Assistants

  Nasir Ahmed Raffick <raffick.n@husky.neu.edu>

  Kairavi Mehta <mehta.kai@husky.neu.edu>

  Alekhya Kotamarthi <kotamarthi.a@husky.neu.edu>
- Talk with your classmates

  Start forming teams (1 – 4 students in a team)

Outline

- Prerequisites
- Course Description
- Textbook
- Grading
- Class Schedule
- Assignments
- Team Project
- Office Hours
- Piazza
- Academic Honesty
- Introduction to Databases
- Designing Databases
- Assignment #1
- Project Topic

Prerequisites

- Experience in Java
- Data Structures
- One year of college calculus
- Discrete Mathematics

Course Description

Introduces relational database management systems as a class of software systems. Prepares students to be sophisticated users of database management systems. Covers design theory, query language, and performance/tuning issues. Topics include relational algebra, SQL, stored procedures, user-defined functions, cursors, embedded SQL programs, client-server interfaces, entity-relationship diagrams, normalization, B-trees, concurrency, transactions, database security, constraints, object-relational DBMSs, and specialized engines such as spatial, text, XML conversion, and time series. Includes exercises using a commercial relational or object-relational database management system.

Textbook

Database Design and Implementation

by Edward Sciore

John Wiley, October 2008

ISBN: 0471757160

ISBN-13: 9780471757160

There will be a reading assignment for each class.

Course Organization and Grading

- 5 Individual Assignments (30%)
- Team Project (20%)
- Mid-Term Exam (20%) 18 October 2016 1.5 hours
- Final Exam (30%) 13 December 2 hours
  - The Mid-Term and Final exams will be an open-book/open-notes exams.
  - Laptops are permitted at the exams with prior approval.
- Submit deliverables to Blackboard.
  - The due time is 11:59pm, the default for Blackboard.
- The grade will be reduced for the following:
  - Late assignment (1 point out of 100 for each hour)
  - Late Mid-Term Exam (1 point out of 100 for each minute)
  - Late Final Exam (1 point out of 100 for each minute)
- Extensions
  - Individual extensions are never given (except certain medical reasons)
  - Class extensions will be given for good reasons that are discussed in class

Grading Scale

Numerical Grade	Letter Grade
93.333-100	A
90-93.332	A-
86.667-89.999	B+
83.333-86.666	B
80-83.332	B-
76.667-79.999	C+
73.333-76.666	C
70-73.333	C-

Class Schedule

- Reading assignments and web links
- Mostly classes will consist of going over example problems related to the assignments.
- See the posted syllabus for the details.

Schedule

- 9/13 Overview of database systems, the relational model and UML design
- 9/20 Database design and translating UML to the relational model
- 9/27 Relational Operators and SQL queries
- 10/4 Advanced SQL queries, updates and views
- 10/11 Integrity, security, triggers and stored procedures
- 10/18 Mid-Term Exam
- 10/25 Index design principles and content management
- 11/1 JDBC database programming
- 11/8 XML, JSON and storage devices
- 11/15 Concurrency control and metadata management
- 11/22 Object-relational mapping
- 11/29 Log, buffer and recovery management
- 12/6 Query processing and optimization
- 12/13 Final exam

Assignments

- Assignments due on Mondays
- Most of them have already been posted
- 9/26 Basic database design and translating from UML to the relational model
- 10/10 SQL queries
- 10/24 Integrity, security, index design and content management
- 11/7 Database programming
- 11/21 XML, JSON, XPath, and concurrency

Team Project

- Teams consist of 1-4 students
- Students can leave a team but not join one
  - If a student leaves, the project scope will be reduced.
- Requirements
  - Specified format
  - Size criteria
- Design
  - UML class diagram
  - Size criteria
- Implementation
- Presentation
  - Specified format
  - Grading criteria
- Report
  - Specified format
  - Grading criteria
- Some teams will give their presentations in class
  - Voluntary

Project Deadlines and Deliverables

- Group formation (No deadline)
  - Deliverables: Members of the group and an optional group name
- Project topic (9/19) 10%
  - Deliverables: Purpose and objectives
- Project requirements (10/3) 20%
  - Deliverables: Use cases and use case descriptions
- Project design (10/31) 20%
  - Deliverables: Varies with project, but will always include a UML class diagram and may include other diagrams
- Project implementation (11/28) 30%
  - Deliverables: Source code
- Project report (No deadline) 20%
  - Deliverables: Document (in any format) and slide presentation

Office Hours

- Monday 3:00-4:00pm and Thursday 1:30-2:30PM
  - or by appointment
- 342 WVH
- Starting 9 January 2017
- Ending 13 April 2017
- There will be no office hours on the following days:
  16 January; 20 February; 6, 9, 27 and 30 March 2017
- The TAs will have office hours
  - See posting on Piazza

Piazza

- Discussions will be on Piazza
- Please use proper etiquette
  - Each posting on a single question
  - Followups should be on the same topic
- Announcements for the course
  - New materials
  - Useful tools
  - Extensions
- Piazza announcements are a required part of the course

Required Software

- SQL Standard Relational Database System
  - Recommended: MySQL
- Java compiler and run-time
- Optional: An IDE such as Eclipse
- Optional: UML drawing tool

Academic Honesty and Integrity

- Separate presentation

Introduction to Databases

What is a database?

- A database is an organized collection of persistent structured data
- A database management system (DBMS) is a software system that manages databases independently of applications that use the databases.
- The DBMS industry is very large and growing
- The Relational DBMS is a major part of the DBMS industry

Files versus Databases

- Files
  - Structure is defined by one application
  - Can only be opened and updated by one application at a time
  - Design is ad hoc
  - Administration is ad hoc
  - Nearly always accessed sequentially so performance is poor unless data is copied into main memory
  - Failure protection is coarse, mostly using daily backups
  - Security is coarse; only a file as a whole is protected
- Databases
  - Structure is independent of any application
  - Can be opened and updated by thousands or millions of applications at the same time
  - Systematic and well developed design tools and methodologies
  - Highly developed administration tools
  - High performance random access with indexes
  - Well developed fine-grained failure and security protection

Using Databases

- Install DBMS
- Design databases
- Develop programs to access and update the databases
- Administer the database
  - Grant and revoke permissions
  - Performance tuning
  - Failure recovery policies

Designing Databases

- Best approach for developing programs
  - Design with high-level modeling language (UML)
  - Implement with high-level programming language (Java, C++, C#, ...)
  - Translate (compile) to low-level language (assembly language)
  - Execute on run-time environment (JRE, CLR, ...)
- Best approach for developing databases
  - Design with high-level modeling language (UML)
  - Translate to low-level language (SQL, ER model)
  - Execute on run-time environment (DBMS)
- Both approaches start with UML
  - Improves compatibility between programs and databases

Designing Databases

Chemistry Design Problem

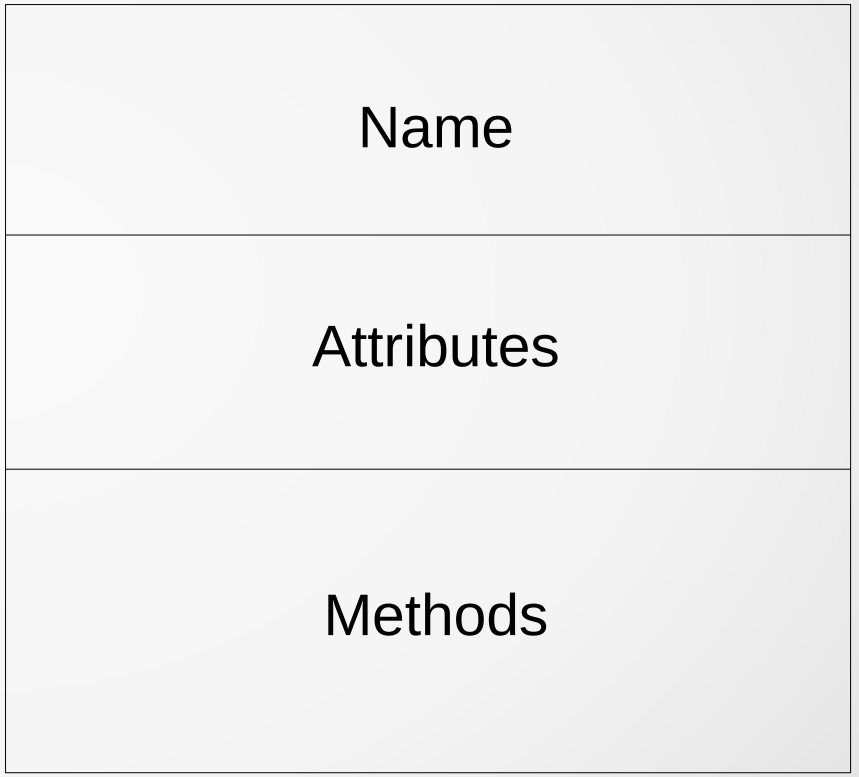
In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

Object-Orientation Fundamentals

- An *object*
  - has object identity
  - remains the same even when its properties change
  - Examples: person, company, tree
- A *value*
  - has no object identity
  - changing a value produces a different value with no connection to the previous value
  - Examples: 5, 3:30pm, “Hello, world!”

Class

- A collection of *objects*
- Each class has a name
  - It should be a singular noun or noun phrase
  - Abstract classes have italicized names
- Attributes are the data values of each object
- Methods define the behavior
  - Not used in database design

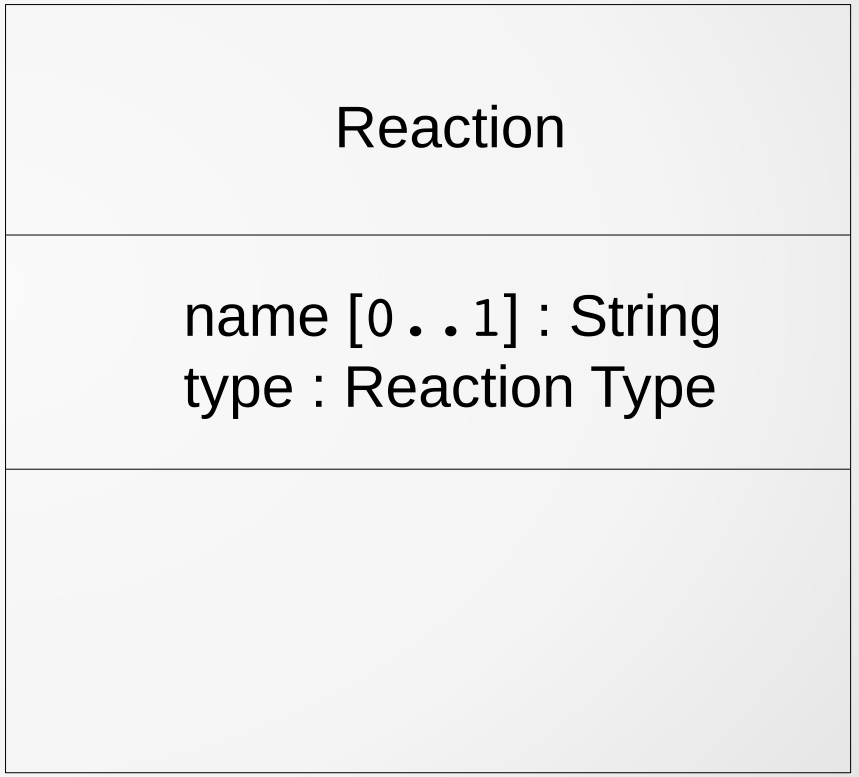


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic weight and an atomic weight. The atomic number is a positive integer, and the atomic number is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. **A reaction may have a name and has a type.** The reaction types are synthesis, decomposition, displacement and combustion.

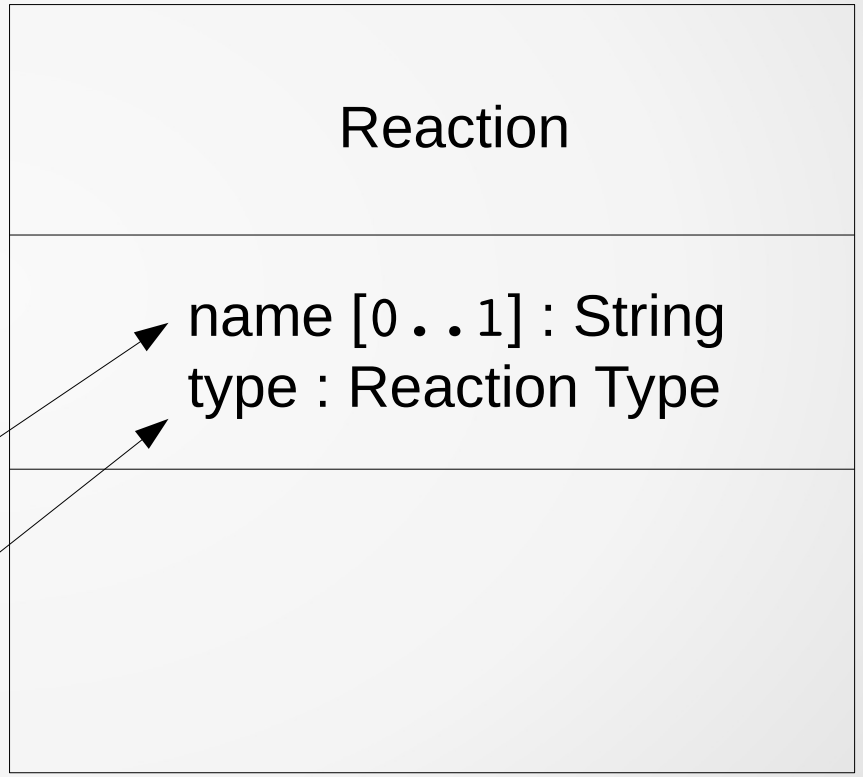
Attribute

- An attribute has a name, a type and a multiplicity
- Built-in types include String, Integer, Double, Boolean, Date, Time, ...
- A type can be
  - Another class
  - The same class
  - An enumeration
  - A datatype
- Multiplicity is on the next slide



Multiplicity

- Defines how many values an attribute can have
- Written [m..n]
  - m is the minimum number
  - n is the maximum number
  - * means unlimited
- For an attribute, the default is [1..1]
- The most common are:
  - [0..1] At most one value (optional)
  - [1..1] Exactly one value (mandatory)
  - [0..*] Any number of values
  - [1..*] At least one value

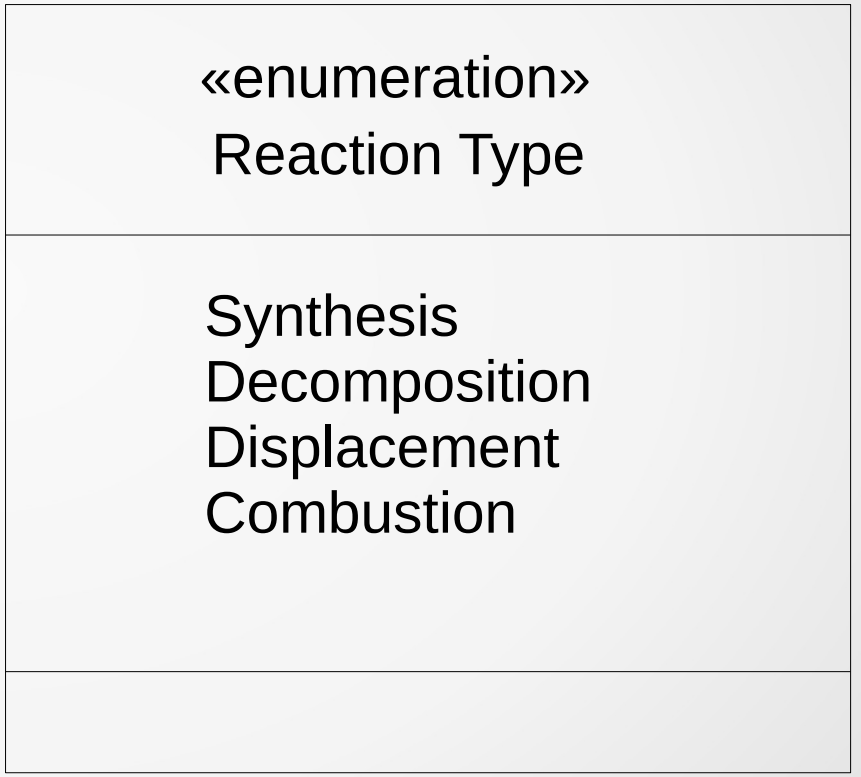


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. **The reaction types are synthesis, decomposition, displacement and combustion.**

Enumeration

- Finite (usually small) number of instances
- Specified as a class with the Enumeration *stereotype*
  - *Annotation* in Java
  - *Attribute* in C#
- Instances are called *literals*

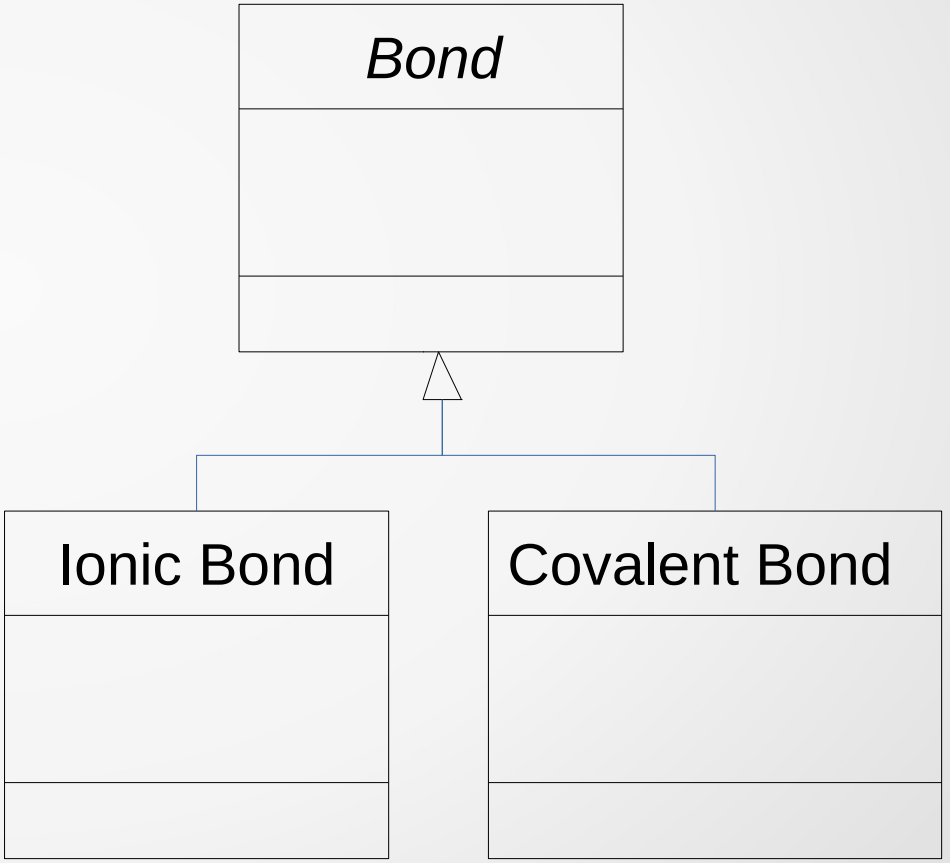


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic weight and an atomic weight. The atomic number is a positive integer, and the atomic number is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. **A bond may be either covalent or ionic.** *A covalent bond has a polarity.* For example, water has a polarity of 1.85. *An ionic bond has a charge.* A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

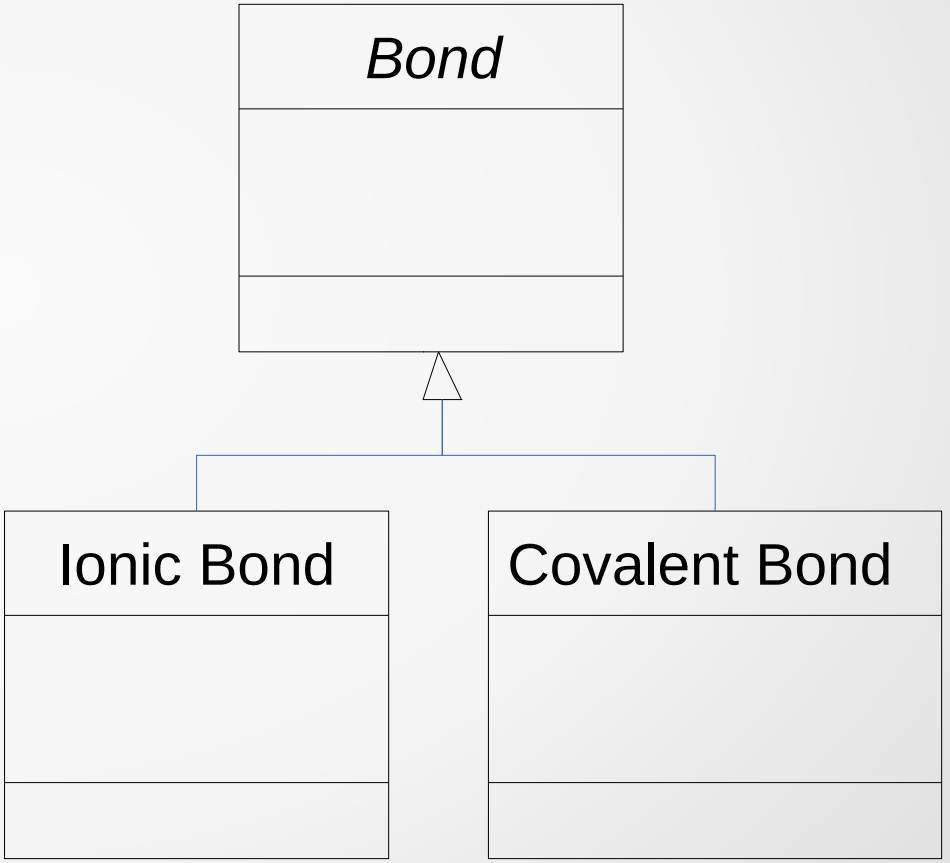
Specialization/Generalization

- Other names
  - Subclass/Superclass
  - Inheritance
  - Derivation
- Since all bonds are ionic or covalent, the Bond class is abstract.
  - However, this does not have to be specified in a data model
- Specified in a design in various ways such as “is a”, “is a kind of”, ...
  - However, beware. A kind could be an attribute!
  - You need additional requirements before using subclasses



Specialization/Generalization

- A subclass is related to the superclass by the subset relationship
  - The set of ionic bonds is a subset of the set of bond.
- There is a way in UML to say that these two subclasses are disjoint, but it is not normally used in data design

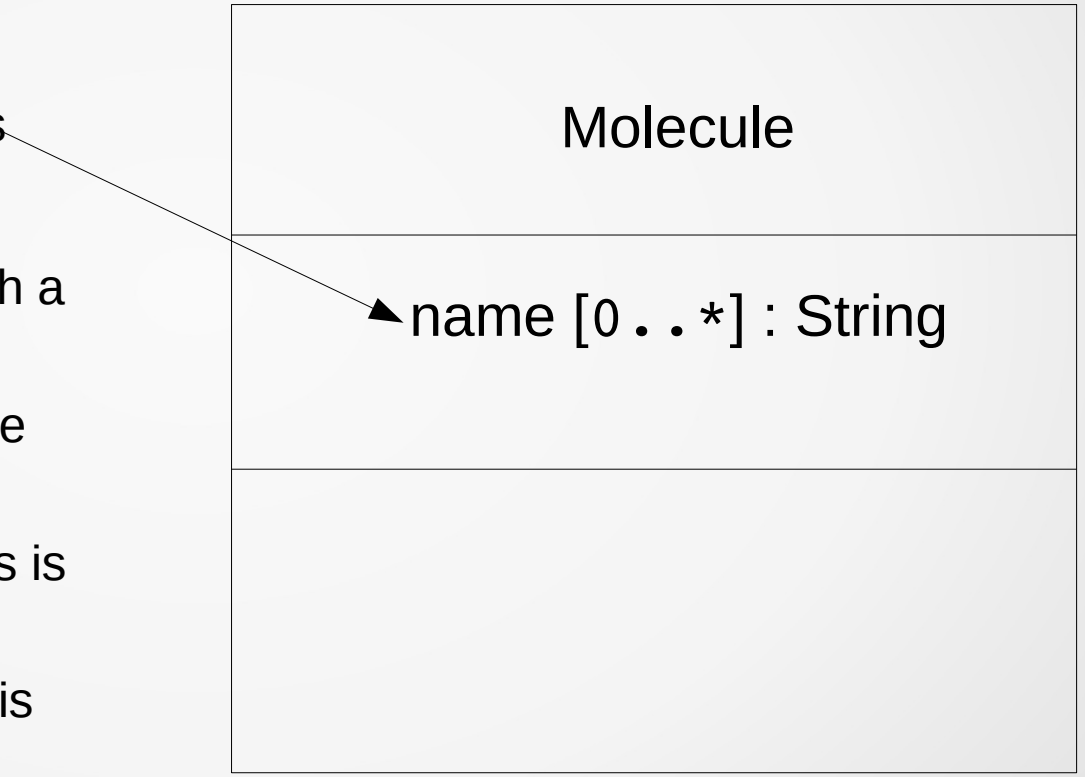


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. **A molecule may have names.** For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

Multivalued Attribute

- An attribute that can have a set of values
- The most common are:
  - [0..*] Any number of values
  - [1..*] At least one value
- Indicated by using the plural with a singular
  - A molecule *may* (or *can*) have names is [0..*]
  - A molecule *must* have names is [1..*]
  - Molecules may have names is ambiguous

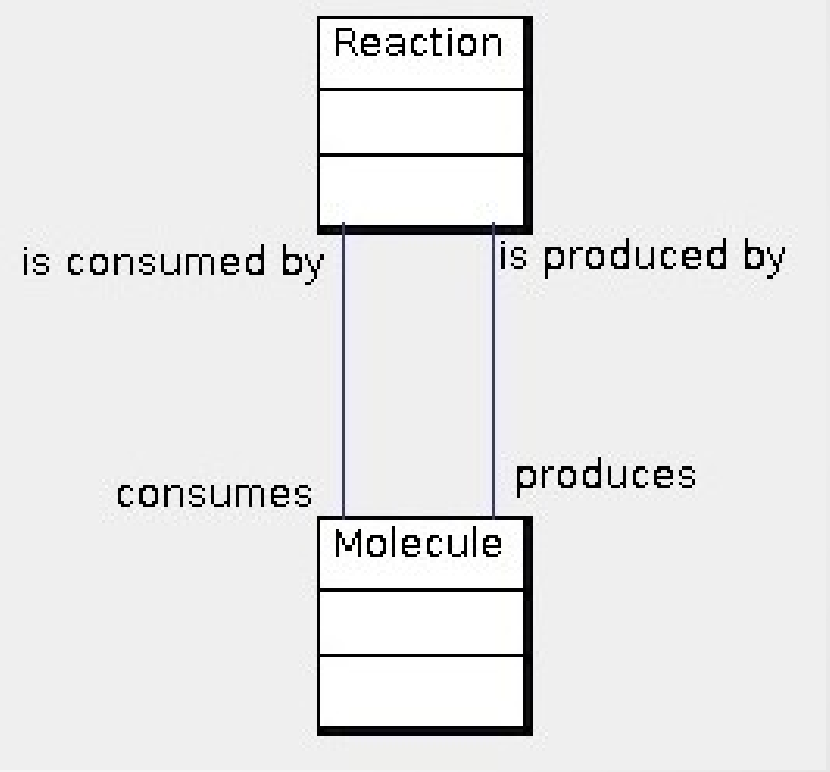


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. **A reaction consumes one or more molecules and produces one or more molecules.** For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

Association

- Other names
  - Relationship
  - Connection
- Specified with a line between the two classes
- Each association has two names (called *roles*)
- The roles are verbs or verb phrases



Association

- The roles should allow one to form a sentence

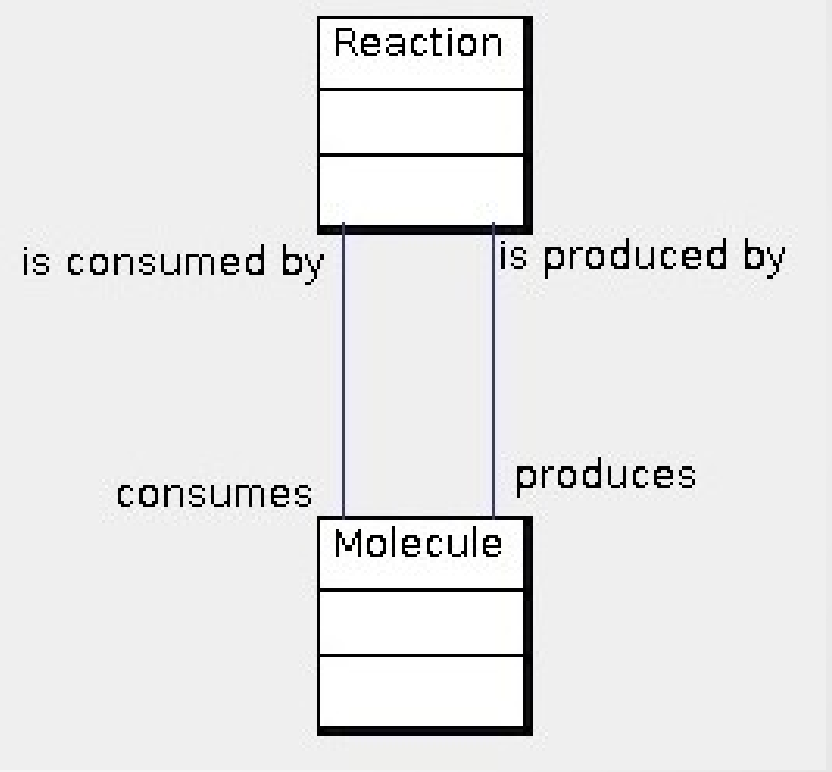
  “A reaction produces molecules”

  “A molecule is produced by a reaction”

- Note that the role is attached to the object of the sentence:

  Subject (Verb Object)

- An association can have a name as well as roles
  - Association roles are required
  - Association names are optional



Association

- An association is a collection of *links*
  - A link is a labeled ordered pair
- The water formation reaction (WFO)

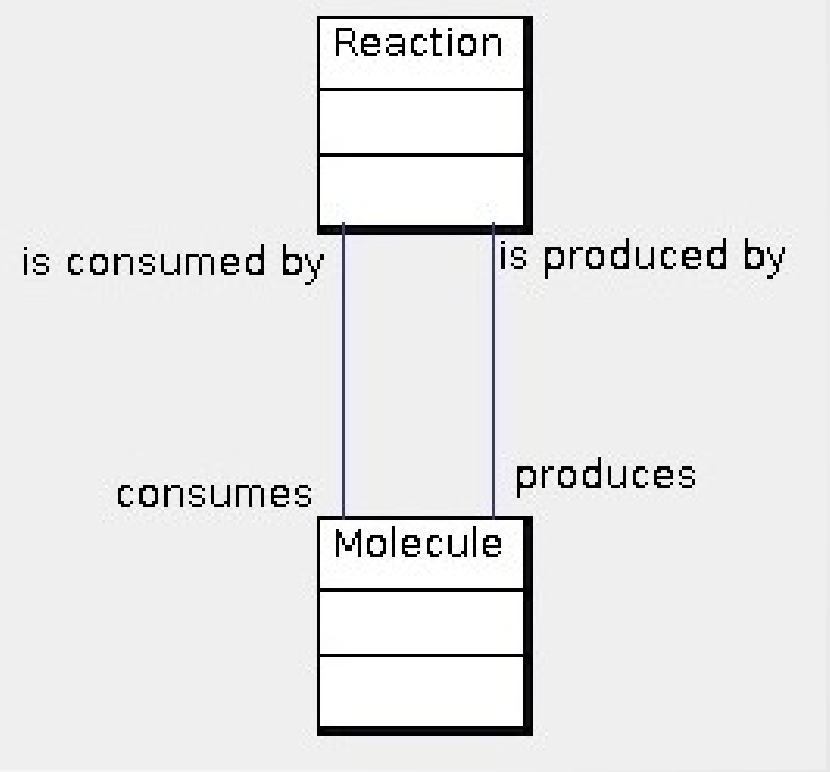
  consumes H_2 and O_2

  produces H_2O
- There are three links:

  WFO – produces → H_2O

  WFO – consumes → H_2

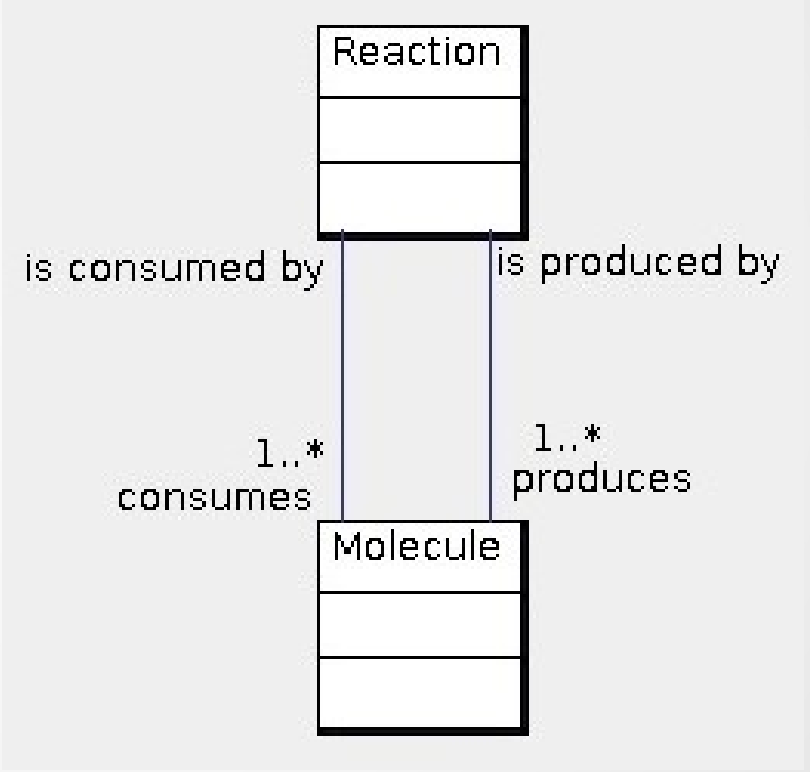
  WFO – consumes → O_2



Association Multiplicities

- Associations have multiplicities
- Each end (role) has a multiplicity
- Both multiplicities *must* be specified
- There are no defaults for ordinary associations
- Note where the multiplicity is placed:

  “A reaction produces one or more molecules”



Association Multiplicities

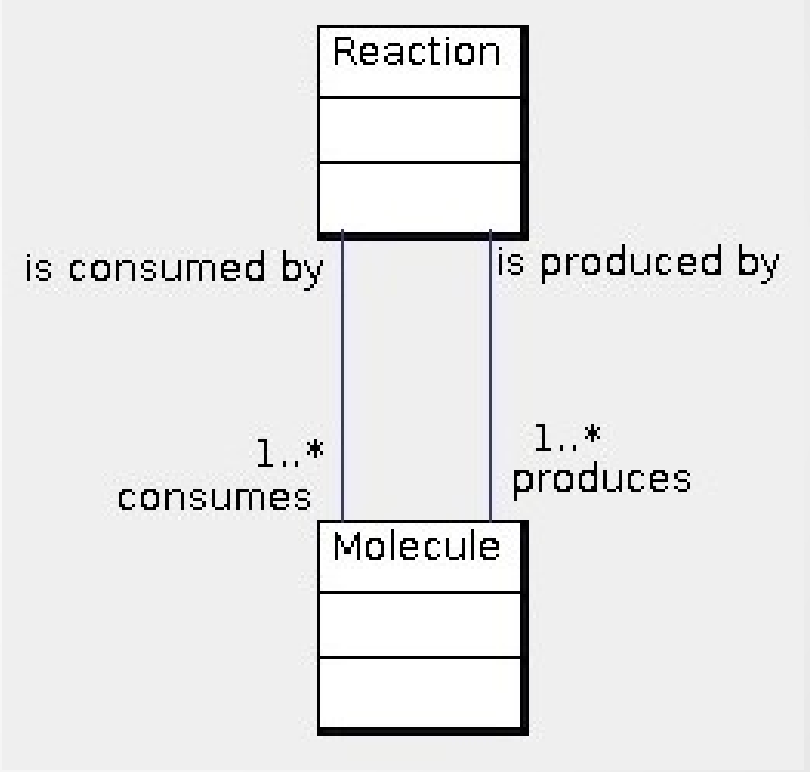
- For example,

  WFO – produces → H_2O

  WFO – consumes → H_2

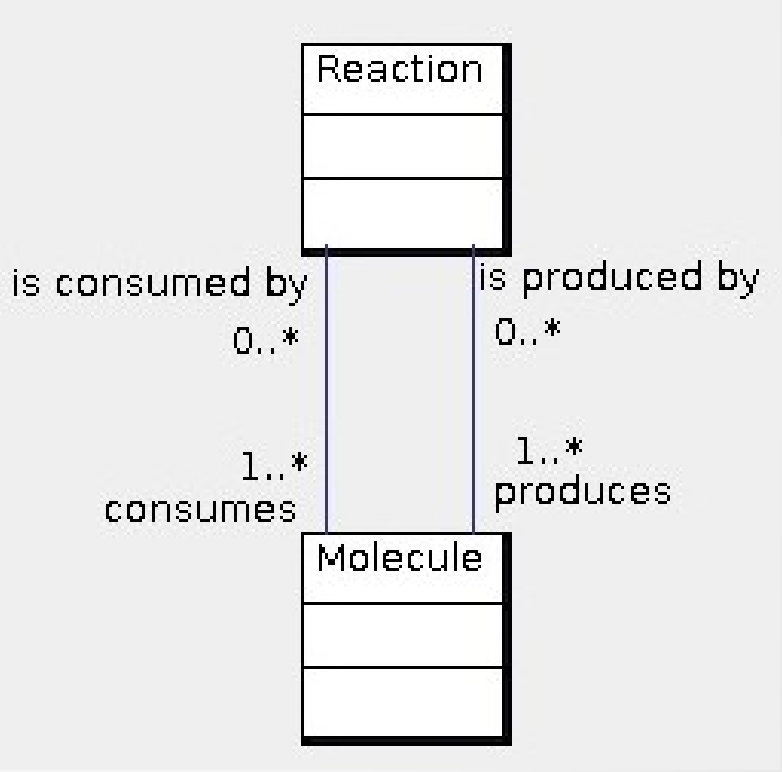
  WFO – consumes → O_2

- WFO consumes 2 molecules and produces 1 molecule



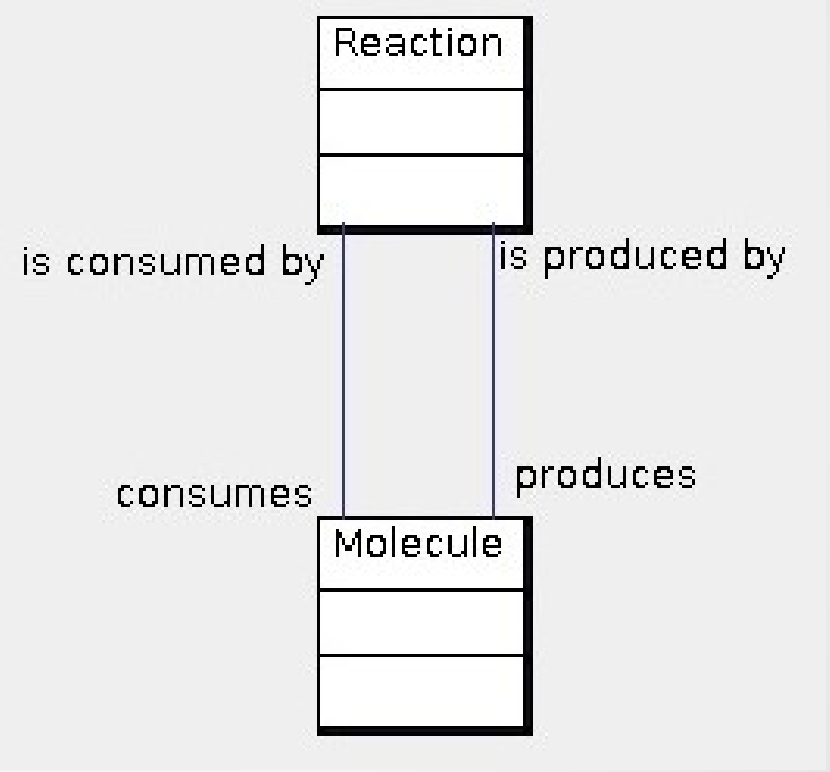
Association Multiplicities

- The multiplicity will usually only be specified in the requirements document on only one end.
  - It does not say “a molecule can be produced by more than one reaction”
- In such a case, the multiplicity will normally be [`0..*`]
- If you are in doubt about this, then ask
  - This is a legitimate question: “Can a molecule be produced by more than one reaction?”
  - This is not an acceptable question: “Is the multiplicity for is consumed by [`0..*`]?”



Objects and Values

- Reactions and molecules are objects
  - WFO is an object
  - H_2O, H_2 and O_2 are objects
- Links are values
  - WFO – produces → H_2O
  - WFO – consumes → H_2
  - WFO – consumes → O_2



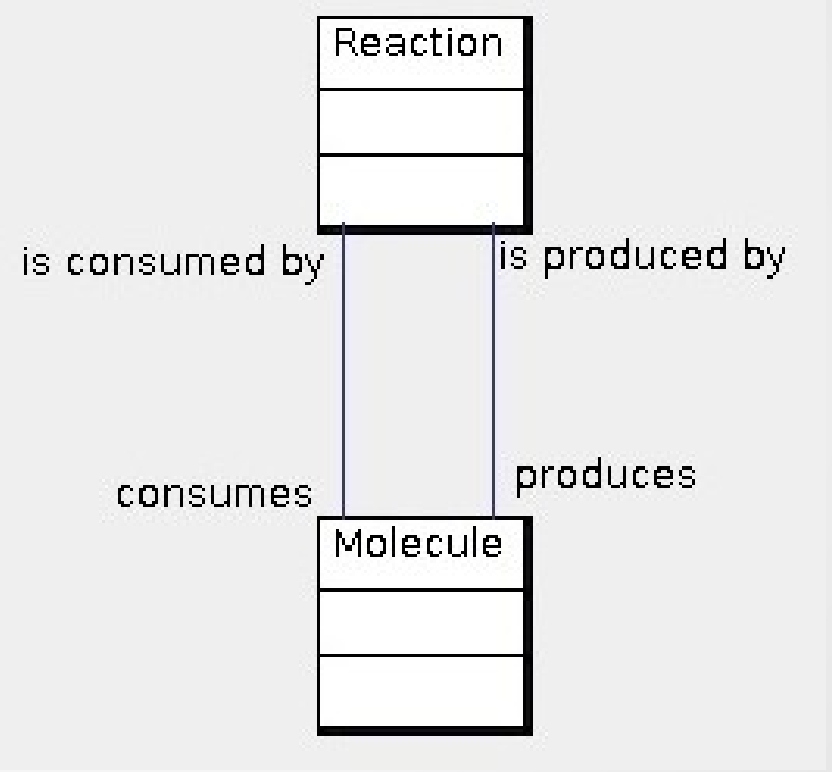
Objects versus Values

- Does it matter that

  WFO is an object, but

  WFO – produces → H_2O is a value?

- Yes, it does!
- Either WFO produces water or WFO does not produce water. WFO cannot produce water more than once.
- If a link was an object then WFO could produce water any number of times.

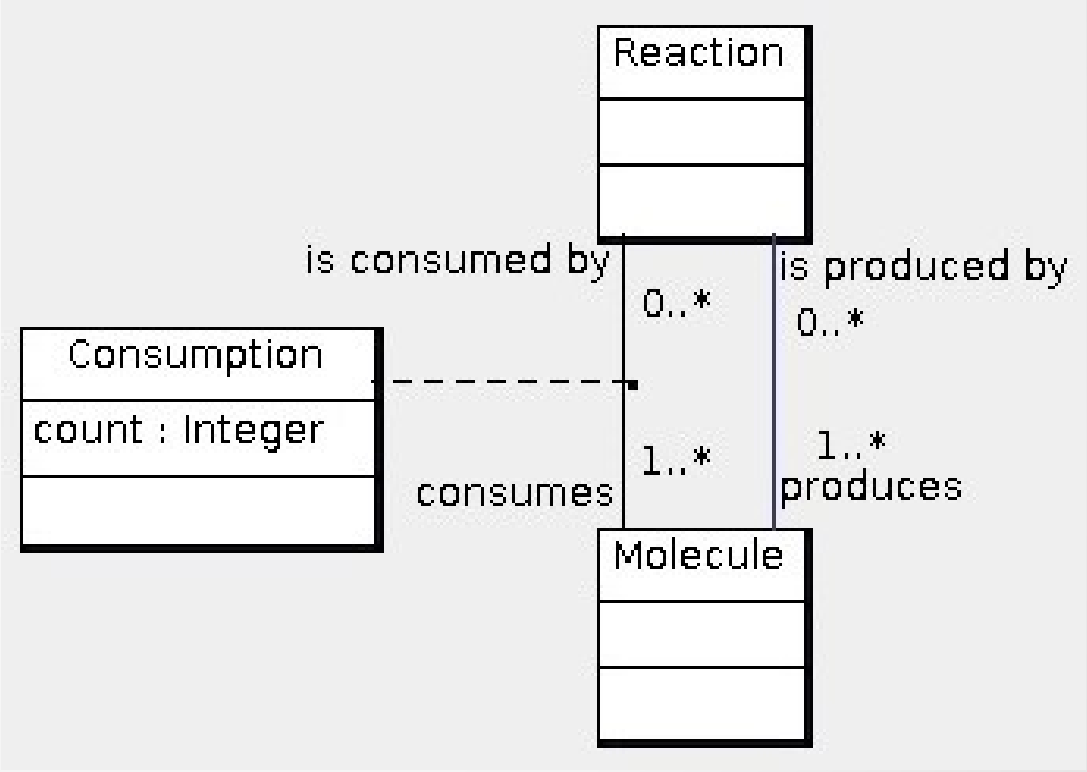


Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called "water." An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. *For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O.* **It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced.** A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

Association Class

- Allows one to add attributes to a the links of an association
- It looks like a class but it is still an association
  - Many textbooks and websites get this wrong!
- Association classes can participate in associations

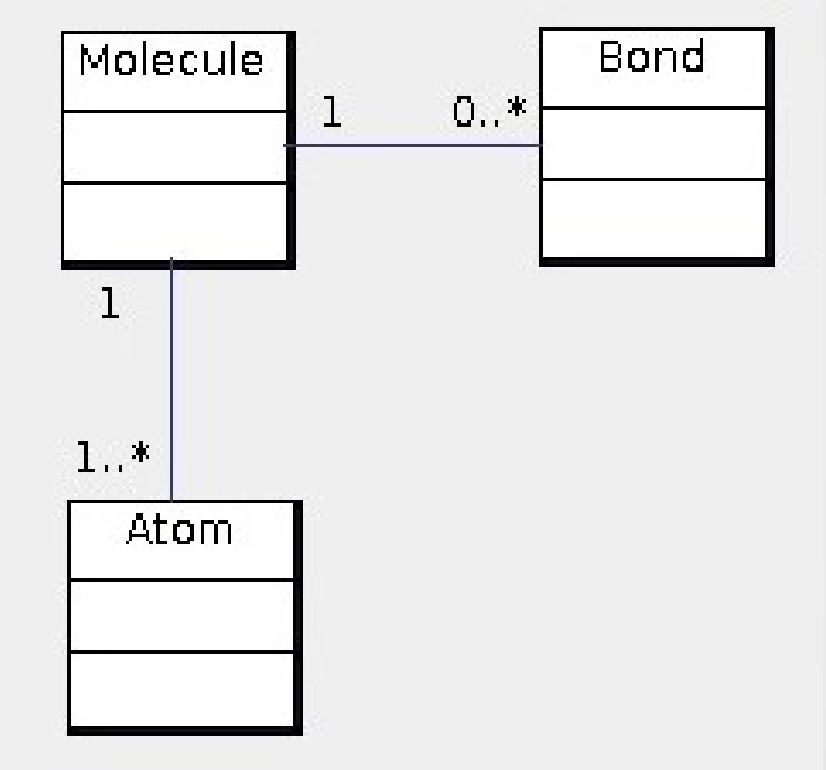


Chemistry Design Problem

In Chemistry, **a molecule consists of at least one atom.** A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. **A molecule may have bonds.** A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.

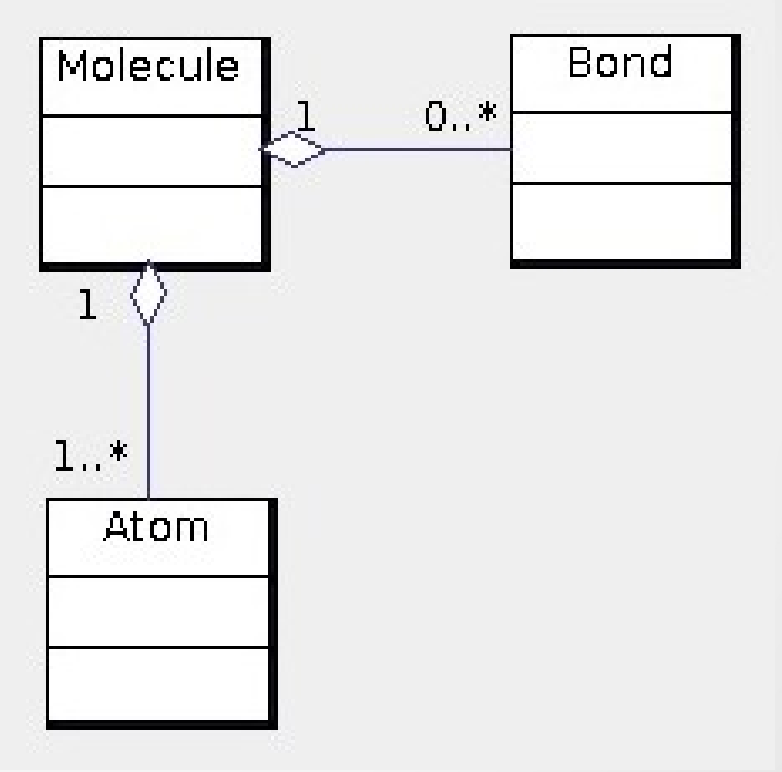
Aggregation

- The design with multiplicities (but not role names yet) is shown here:
- However, a molecule *consists of* atoms. This is a containment relationship.
- An atom is a *part of* a molecule.
- This is called *aggregation*



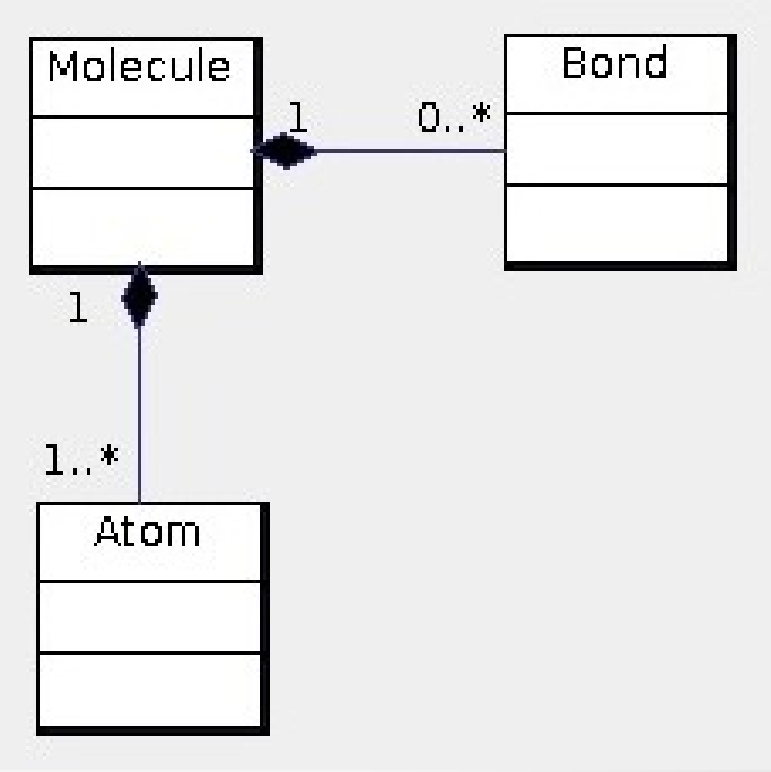
Aggregation

- Aggregation is shown as an open diamond on the container end:
  - Think of the open diamond as the “box” that contains the objects on the other end
- Unlike ordinary associations one does not have to show the role names. The defaults are “contains” and “part of”
- There are default multiplicities: [1..1] and [0..*]



Composition

- An aggregation is an ordinary association but with default multiplicities and role names.
- However, one can say more in this case
  - A molecule is *composed of* atoms and bonds.
  - If the molecule is deleted, then the atoms and bonds are also deleted.
- This is called a *composition*.
  - Shown as a filled in diamond



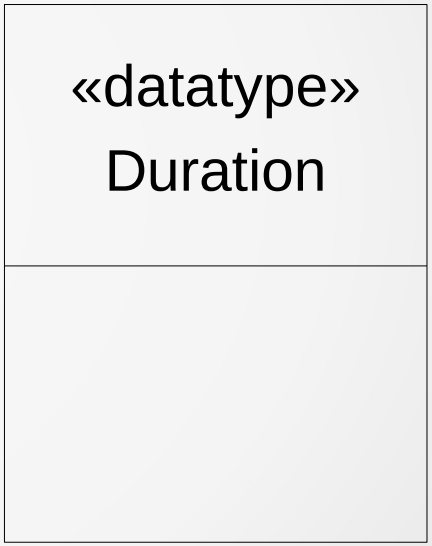
Aggregation versus Composition

- An aggregation is an ordinary association but it has default multiplicities and role names.
- A composition is *not the same* as an ordinary association.
  - A composition has the same defaults as an aggregation
  - A composition adds an additional constraint to an aggregation:
    - Deleting a composite container causes the deletion of the objects that it contains
    - This is called *cascading deletion*

Datatype

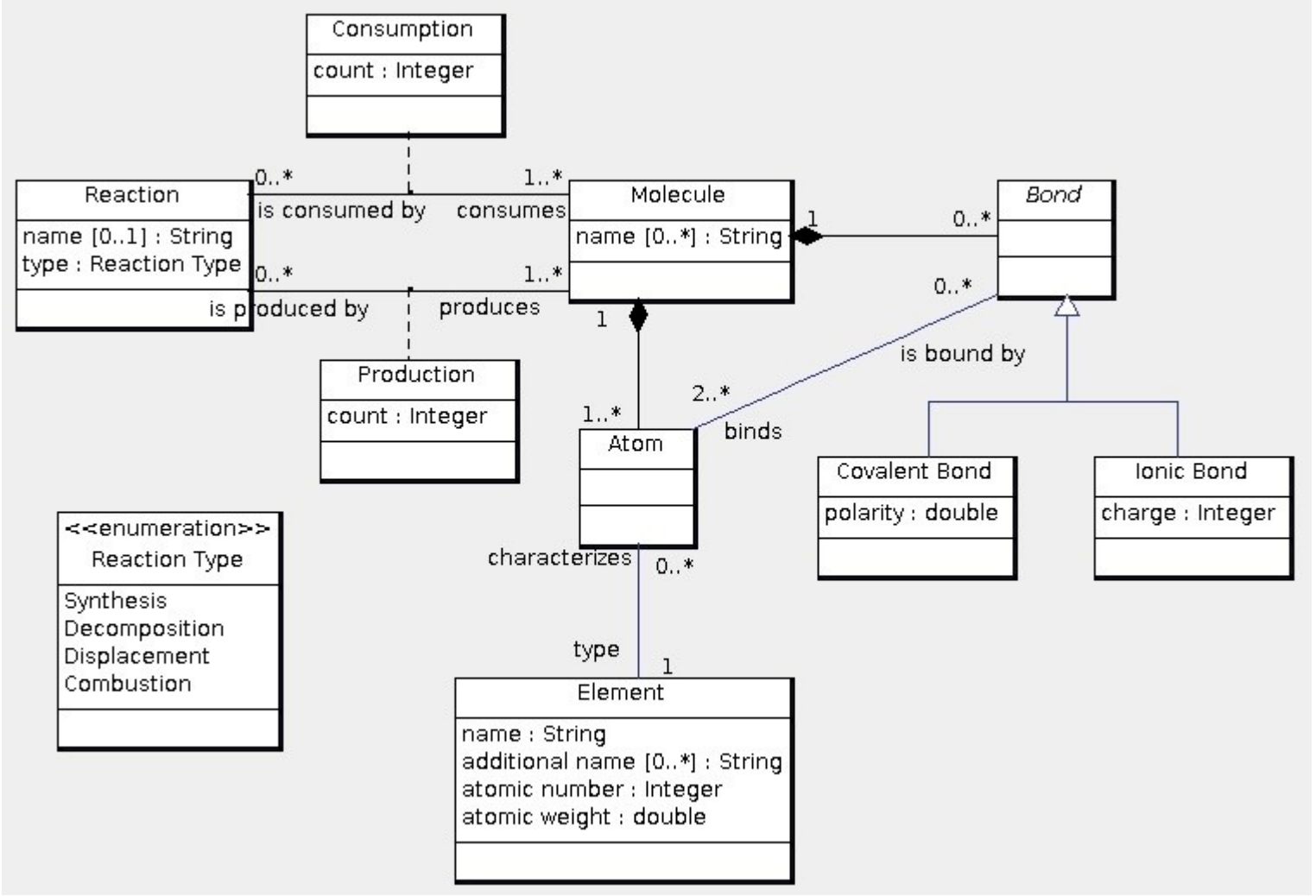
- Specified with a stereotype
- Examples of datatypes and values
  - Double 5.6
  - Date September 13, 2016
  - Time of Day 6:00pm
  - DateTime September 13, 2016 6:00pm
  - Geographic Coordinates

    41°24'12.2"N 2°10'26.5"E
  - Duration 38 minutes
  - Boolean true



Chemistry Design Problem

In Chemistry, a molecule consists of at least one atom. A molecule may have names. For example, H_2O is commonly called “water.” An atom is a type of element. For example, water has 2 atoms whose type is hydrogen and one atom whose type is oxygen. An element has a standard name, and may have additional names. An element has an atomic number and an atomic weight. The atomic number is a positive integer, and the atomic weight is a floating point number. A molecule may have bonds. A bond binds 2 or more atoms. A bond may be either covalent or ionic. A covalent bond has a polarity. For example, water has a polarity of 1.85. An ionic bond has a charge. A charge is a positive integer. A reaction consumes one or more molecules and produces one or more molecules. For example, two molecules of H_2 and one molecule of O_2 react to form two molecules of H_2O. It is necessary to keep track not only of which molecules are consumed and produced but also how many of each molecule is consumed or produced. A reaction may have a name and has a type. The reaction types are synthesis, decomposition, displacement and combustion.



Diplomacy Design Problem

Design a database for the following kinds of data. A diplomat is a person appointed by a state. A diplomat conducts diplomacy with one or more other states. A person has a name and unique email address. A diplomat has a rank. A state has a unique name. The diplomatic ranks are attache, counselor, envoy, minister and ambassador.

Everyone should try this. Write your solution on a piece of paper.

Assignment #1

Project Topics

Requirements

- Due 22 January 2017
- Team Members
  - You can leave a team
  - You cannot join a team after 22 January 2017
- Project Purpose
- Project Description

Project Examples

- **SalesGun** is an idea born from the possibilities that might open up if brands are connected to people, in simple terms the app lets people publicize their preferred products to friends and earn credits while doing it.
- The **Distributed Generation System Repository** is a database that contains several distributed energy generation systems, including
- Renewable generation systems and Traditional (Fossil fuel) generation system
- Online Learning Management System
- **Dine - In** Restaurant Web Service
- The European Championship Gambling Online
- Library Management System

Project Examples

- Disease Prediction from Health Care Assessment System
- **BLOG AWAY**: A BLOGGING WEBSITE
- ONLINE AUCTION SYSTEM
- Online Music Library
- UTILITY MANAGEMENT SYSTEM
- Used Item Exchange Website
- **Convergence** is an online discussion board application which makes it convenient to share information on different topics and help users to get solution

Project Examples

- **DOUBAN SAUCE WEBSITE**: A Real Estate Information Platform
- Online Finance Management (OFM)
- Physical Geography Database System Design
- FIFA World Cup Database Management System
- Foster Care Management
- Hotel Reservation System