Preliminaries...

We need to talk about some terminology.

An entity is any thing that can be distinguished from other things. Each entity is characterized by a set of properties called attributes. The set of values for each entities attributes completely distinguish it from other entities.

If multiple entities have a common set of attributes (not values) we say that they belong to the same entity set.

Attributes can be:

atomic: char* ssno
composite: a set of atomic attributes, phone number is composed of an area code, exchange and number
multi-valued: a class list
derived: a gpa.

An attribute (or set of attributes) is called a key if it uniquely describes an entity.

The set of attributes that describes an Entity Type is called its schema. Each attribute takes values from a specified set called its Domain.

(In OO terminology, entities are objects and entity sets are classes).

Entities (and therefore entity sets) may have relationships to each other.

Several such relations can be identified:

Hierarchy..( ISA ).one entity is a generalization of another, vis-à-vis one is a specialization of another.. examples:

A faculty member "is a" specialization of a university employee.

Student is a generalization of the entity Sets Graduate Student, Undergraduate Student, Visiting Student.

Contains..(HASA). one entity contains another.. example

A class (in the sense of a university) contains students.

Sometimes expressed as "is part of". A student is part of a class.

Uses.. .one entity uses another to accomplish some task.. example

An airline flight uses an Aircraft.
Steps in the database design procedure. These are not necessarily linear, some are done in parallel, some looping can occur.

1. Requirements Analysis
2. Conceptual Design
3. Choose a Data Model (Relational)
4. Map the design to the model
5. Normalization
6. Physical Design.

1. Requirements Analysis: Gather Data.

   Inputs:
   a. User interviews
   b. Analysis of Operating Environment
   c. Analysis Of Processing requirements
   d. Review of past documentation

   Outputs:
   - A document that identifies the data objects (Entities), the relationships between these objects and the rules that govern how information is created and manipulated. The document should also include the scope of the database, its users and its applications.
   - A data dictionary

Tools for this step fall into a category of called Systems Analysis

We will attempt to identify

   - The principle entities that comprise our system
   - Relationships between these entities
   - The Business Rules that specify how the company does business.
   - A set of scenarios that describes the processing required for the enterprise.


   Inputs: The documents of the previous step.
   Output. A global schema that shows all the data and their relationships. In particular, this step should produce a list of entities, attributes relationships and constraints on the system.

Tools: Entity-Relationship Diagrams, UML and many others.
There are two extreme strategies:

a. Centralized scheme design.. all the user requirements are merged into a single set before modeling.

b. View integration.. A schema is developed for each group of users based on their separate requirements. The resulting views are consolidated into a single global design.

Conceptual Design. Part Two.

A conceptual design should also include a complete list of processing requirements and transactions that are required for the enterprise.

Tools: CRC Cards, Storyboarding, Data Flow Diagrams.

3. Choose a database. Choice depends on the type of problem

```
Problem Space

<table>
<thead>
<tr>
<th>msec.</th>
<th>T</th>
<th>X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>OO Db</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td>Rel Db</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (# of entities per transaction)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tx Complexity</td>
</tr>
</tbody>
</table>
```

Step 4. Mapping to the data model

input: output of step 2

output: A set of data structures appropriate to the DBMS chosen in step 3.

// We'll be constructing tables

Step 5. Normalization

input: the output of step 4 plus a set of functional dependencies derived from the Business rules.
output: (In a RDBMS) a set of tables in a specified normal form (Usually 3NF or BCNF)

Step 6: Physical Design

input: All of the above

output: A collection of files which have been structured and indexed so as to maximize the performance of the system.

<DESIGN EXERCISE SCHEDULE>

Getting Started with the Requirements analysis phase....

Begin by examining written documentation and interviews with users...

1. Create a list of entity types by the following steps:
   a. Make a list of the noun phrases that occur.
   b. From that list, identify noun phrases that represent physical objects and coherent abstractions that are important for the functioning of the system. Place these nouns in a tentative data dictionary.
   c. Delete from the list any nouns that represent objects outside the system or are too vague.
   d. Examine the tentative data dictionary and eliminate synonyms and nouns that represent the attributes of other things. Examine the list for homonyms and carefully consider the effect of adjectives.

To create a list of Transactions examine the verb phrases...decide if the verb represent some responsibility of the system.

Next Logical Design

Defn: Given two of Entities types $E_1$, $E_2$, a relationship, $R$, defines a correspondence between entity occurrences...the number of entity types that occur in the relationship is called the arity of the relationship

eg's

$\text{WorksOn}: \text{Employee} \leftrightarrow \text{Project}$

(boyno, accreditation)

(wolff, Math ed Phd)

This is a binary relationship

$\text{ClassMeeting}: \text{Class} \leftrightarrow \text{Room} \leftrightarrow \text{Hour}$

(Cmpt592, RI-102, W 5:20)
is a ternary relationship

Relationships can be recursive

PlaysAgainst: Teams <-> Teams
(MSU, WPC)

Re-occurring relationship among the same entity type is called a relationship type.

Note that relationships may have attributes...a grade is an attribute of the relationship between a class and a student...

Entity Relationship Diagrams (P. Chen 1976) R p358

Entity Types are represented by rectangles
Relationship types by diamonds
attributes by ovals
lines are used to connect the pieces

There are several ways to modify ER diagrams so that they contain more information...