What to build a mansion/house?
==========
What should be done - Dear Naresh, Naveteja and Ashitosh

Location of place / Area / Soil   => HYD
Construction material (Later)

Size of the plot   = 30 x 40 Feet
Requirements   = 2 Bedrooms, 1 kitchen, 1 Hall + Dining
                Attached wash room
                Vastu, GHMC laws
Planning   -> PAPER
      Layout plan
      Structural plan
      Front-Elevation plan
      - Plumbing design
      - Electrical design
      Interior design plan
- What material to use
- Normal construction or Column/Beam structure
- Estimation

==========
Definition of Data Model
~~~~~~~~~~~~~~~~~~~~~~~~
A DATA MODEL is an INTEGRATED COLLECTION of
- CONCEPTS for describing data,
- RELATIONSHIPs between data, and
- CONSTRAINTS on the data
used by the organization/company

A DATA MODEL is a representation of 'real world' objects, events
and their association.

Example: Payroll Application
To generate a Payslip (Which is done by our Payroll Application)
- Describe the data
  * No. of attended
  * EmpID -> Name and gender
  * Department   -> Admin, Acc, Sales and Prod
  * Bank Details -> A/c
  * Salary Amount (Basic Salary) => What type of data? -> NUMERIC
      Min and Max
  * Designation / Grade
  * Overtime / Addon....

==========
Why use Data Modelling?
~~~~~~~~~~~~~~~~~~~~~~
[1] Data Consistency
[2] Scalability
[3] Leverage
    Data model acts as a blue-print
[4] Conciseness
    Effective communication tool
    A PICTURE IS MORE THAN 1000 WORDS
[5] Data Quality

Evolution of Data Modelling
~~~~~~~~~~~~~~~~~~~~~~~~~~
Different type of Flat Files
- CSV (Comma Separated Values)
- XML
- JSON

Relational Database Systems
- Oracle
- MySQL or MariaDB
- Microsoft SQL Server
- SQLite
- PostgreSQL
- DB2
:
:

Features of a Good Data Model
~~~~~~~~~~~~~~~~~~~~~~~~~~~~
* Completeness
* Non-Redundant
* Business Rules
* Communication
* Integration
* Avoid Conflicting objective

==========

OLTP - OnLine Transaction Processing
OLAP - OnLine Analytical Processing (Data Warehouse)

To design your own generic model there two methods:
[1] Bottom-Up Modelling
[2] Top-down Modelling

==========
Characteristic of OLTP
==========
- Application Oriented
- Detailed Data
- Data will be current & Up to Date
- Isolated data
- Repetitive access
  i.e. Application will be accessed again & again
- Clerical User
The model used is ER-Model, Object-Oriented Model
                  ER==>Entity Relationship

Characteristics of OLAP
==========
- Subject oriented
- Business analysis   - Large volumes accessed at a time
- Summarized & refined   (Big Data)
- Time varying   - Mostly read (batch updates)
- Non-volatile   - Redundancy is accepted.
- Ad-hoc access   - Data size is 100GB to TB/PB

The model used id Dimension Model (Star Modelling)

Standardized File Format
~~~~~~~~~~~~~~~~~~~~~~~~
CSV -> Comma Separated Values files
XML ->
JSON -> JavaScript Object Notation

These are all PLAIN TEXT FILES.

==========

ER Model
~~~~~~~~
- High-level conceptual model
- Represents data in terms of - ENTITIES, ATTRIBUTES and RELATIONSHIPs

ENTITIES
~~~~~~~~

Employee		Person
EmpID	unique	Aadhar No / SSN
Name		Name
Date of Joining		Date of Birth
Desgination		Gender
Salary		Mobile Number
Contact Number	unique	Address
Email ID	unique	Email ID
Account No.	unique	
:		
Manager No		

Student	Courses			
Reg.No	Course ID	c101	c102	c103
Name	Course Name	B.Sc	BBA	BCA
Date of Birth	Duration	3	3	3
Course ID	Fees	42500	55600	72500

Strong v/s Weak
~~~~~~~~~~~~~~~~
STRONG (Independent) Entity - NOT EXISTENCE dependent on some other entity
      Employee   Account
WEAK (Dependent) Entity - EXISTENCE dependent on some other entity
      Dependent(s)   Nominee

Types of Attributes
~~~~~~~~~~~~~~~~~~~

Simple	Composed of single component	
Composite	Composed of multiple components	Ex: Address
Multi-Value	Holds multiple values	Ex: Qualification,
Derived	Derivable from a related attribute	Ex: HRA, PF

There could be multiple CANDIDATE KEYs for an ENTITY.
Of those multiple CANDIDATE KEYs, the KEY which we choose to
 uniquely identify an ENTITY is called the PRIMARY KEY(PK)
~~~~~~~~~~~~~~~~~~~

In the INVENTORY entity, observe the following data:
INVENTORY ->Store Name, Part No. and Quantity

Store1	P1	50
Store1	P3	20
Store2	P2	100
Store2	P1	30

With the above data, if we need to uniquely identify the INVENTORY item, we can combine the 'Store Name' and 'Part No.' to uniquely identify a row/record.

In such an event, the 'Store Name' and 'Part No.' together is called the COMPOSITE PRIMARY KEY

COMPOSITE PRIMARY KEY - Consist of two or more attributes.

==========

Relationship

A RELATIONSHIP represents an association between entities.

Different Entity Relationships

1:1 (Binary Relationship)
- A college has a Principal
- A State has a CM
- The nation has a PM

1:N (One to Many)
- A class has students
- A college has courses

N:M (Many to Many)
- Courses are taken by Students
- Faculties handle Subjects

Degree of Relationship

Is the number of entities associated with the relationship.

[a] Unary
- Recursive Relationship
- In the entity 'Employee' we may have a 'Manager No.' attribute.
- But who is a Manager? - Of course an employee only.
- Ex: When we map an employee to a manager, then such a relationship is termed as Unary relationship
- (SELF-JOIN)

[b] Binary
- Relationship between two entities.
- Most common
- Ex: DEPARTMENT has EMPLOYEES

[c] Ternary
- Relationship among instances of THREE entity types.
- Ex: SALESPERSON sells PRODUCT to CUSTOMER

Whenever, we have a TERNARY relationship an ASSOCIATIVE Entity needs to be created.

i.e. Intersection data in an N:M relationships can be stored in a special entity called the Associative Entity

Associative Entity is also called as Relationship Entity.

==========

Notations used in ER Model

==========

Entity	SINGLE LINE Rectangle
Weak Entity	DOUBLE LINE Rectangle
Relationship	Rhombus / Diamond
Attribute	SINGLE LINE Oval
Key Attribute	SINGLE LINE Oval UNDERLINED
Multivalued Attribute	DOUBLE LINE Oval
Composite Attribute	Multiple branched Ovals
Derived Attribute	DOTTED LINE Oval

Specialization

~~~~~~~~~~~~~~

The process of MAXIMISING the difference between the members of an entity by IDENTIFYING DISTINGUISH Characteristics.

Example:

Staff -> Manager, Secretary, Sales Personnel

manages to whom

Generalization

~~~~~~~~~~~~~~

Is a process of MINIMIZING the difference between entities by identifying COMMON Characteristics/Features/Attributes

Generalization hierarchy is a form of abstraction that specify two or more entities sharing the common attributes, that be generalized into a higher level entity called the SUPER TYPE or GNERIC Entity.

The lower level entities are called SUB-TYPE or CATEGORIES.

Example:

Person -> Name, age, dob...

======

Women	Men	Employee	Student
=====	===	========	=======
		EmpID	Reg.No
		Salary	Course
		Designation	:
		:	

Total generalization is represented by SOLID ARROW.

Aggregation

~~~~~~~~~~~

It is a relationship between the WHOLE and its PARTS.

Described as "part-of" relationship
~~~~~~~~~~~

Example: Software Product
          ================
            | |
      Program/App   User Guide

i.e. We say, an 'User Guide' is part of the Software product.

Example:
          Car

      Tyres Engine   ....

NOTE: The sub-type entity exists if and only if the super-type entity
      exists.

==
Normalization
=============
Is a process of DESIGNING Relational Database tables.

Normalization is USED:
- to MINIMIZE Data Redundancy (duplication of information)
- to SAFE GAURD database against insertion/deletion/updation problems.
- to SAFE GAURD database against DATA INCONSISTENCies

It is a process of EFFICIENTLY ORGANIZING data in a DB.

We say a DB is NORMALIZED if it satisfies frist three normal forms.

Functional Dependency
~~~~~~~~~~~~~~~~~~~~
Let us suppose we have two columns A and B, such that, for given value
of column A there is a single value of column B associated with it.

Then column B is said to be functionally dependent on column A.
A -> B

Example:
In an 'Employee' entity we may have 'Salary' column and 'HRA' column.
(HRA = House Rent Allowance). The HRA is 40% of the salary. Then we
say HRA is functionally dependent on Salary.
      Salary   -> HRA
      EmpID -> EmpName

NOTE: The higher the NF applicable to the table, the less vulnerable
      it is to inconsistencies.

==
1 NF
====
   Only atomic values and no repeating groups
2 NF
====
   Every non-key attribute is fully functional dependent on the WHOLE
   (COMPLETE) key.
   i.e. Eliminate PARTIAL dependencies
~~~~~~~~~~~~~~~~~~~~

3 NF
====
   Every non-key attribute is non-transitive dependent on the PK
   i.e. Eliminate TRANSITIVE dependencies

##
What is Physical Database Design?
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
- Converting entities to tables
- Converting attributes to columns
- Converting relationship to Foreign keys
- Defining Constraints

Purpose - to OPTIMIZE PERFORMANCE as much as possible

Inputs to Database Design
~~~~~~~~~~~~~~~~~~~~~~~~~
Along with Logical Data Model, the DATABASE DESIGNER requires the
following:
- The process model
- The mapping
- Non-structural data requirement
- Performance requirements
- The target RDBMS
  Like - Oracle, MySQL, DB2, Microsoft SQL Server etc.
- Disk space requirement
- Availability of skilled resources

Designing OLTP DB
~~~~~~~~~~~~~~~~
Translating ER data objects into tables.
Transforming attributes into columns and
  relationships into joins
Determine PK, FK, Unique key etc.
Normalize the data model.
Choosing the data type for columns

Implement the Relational Data Model involving the following:
- Creating the database & its objects
- Populating the database
  (Filling the database table(s) with ACTUAL DATA)
~~~~~~~~~~~~~~~~