



SAPIENZA
UNIVERSITÀ DI ROMA

DATA MANAGEMENT

Assignment 1 - 2

Students:

- **Andrea Potì 2008416**
- **Giulio D'Erasmus 1859130**

The Dataset: Human Resource

HR: Analytics

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked
Over18	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StandardHours	StockOptionLevel
TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager

HR: Data Generator

Emp ID	Name Prefix	First Name	Middle Initial	Last Name	Gender	E Mail
Father's Name	Mother's Name	Mother's Maiden Name	Date of Birth	Time of Birth	Age in Yrs.	Weight in Kgs.
Date of Joining	Quarter of Joining	Half of Joining	Year of Joining	Month of Joining	Month Name of Joining	Short Month
Day of Joining	DOW of Joining	Short DOW	Age in Company (Years)	Salary	Last % Hike	SSN
Phone No.	Place Name	County	City	State	Zip	Region
User Name	Password					

GOALS:

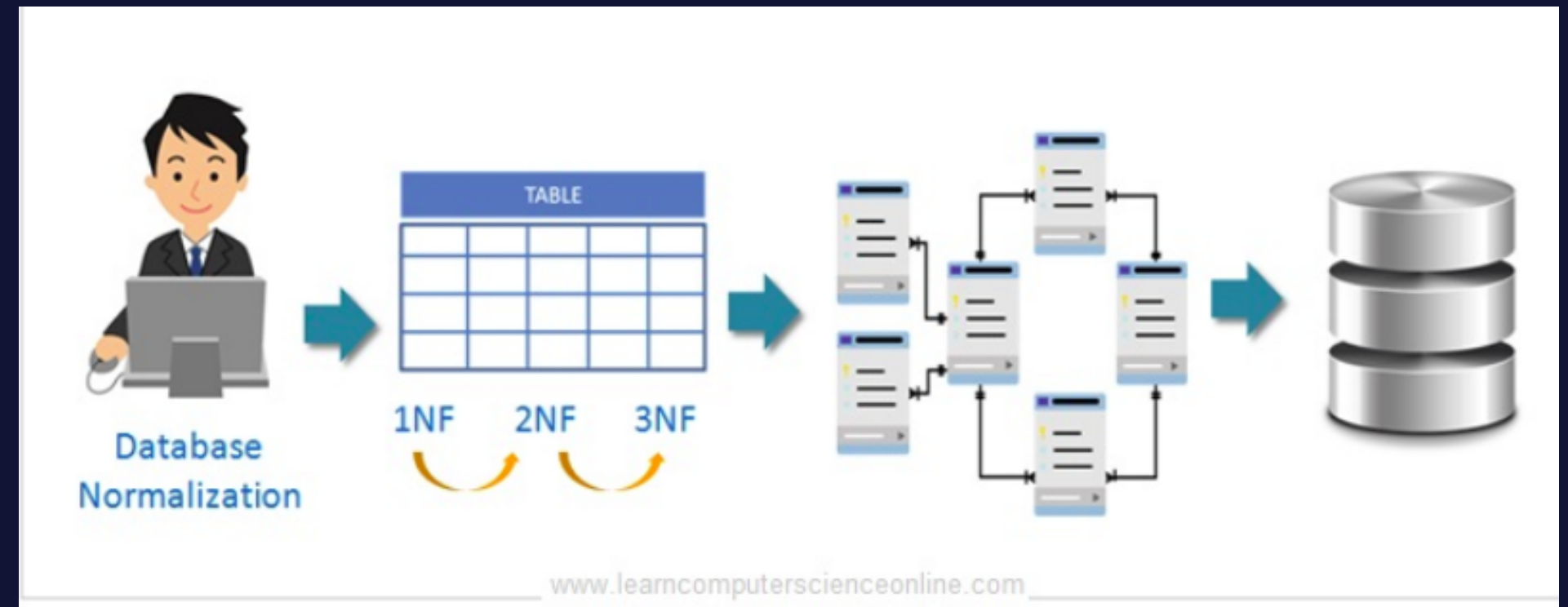
- Avoid redundancy of data:

NORMALIZATION

- Constrain:

- Primary key; 
- Foreign key. 

- Optimized data types
- Data preprocessing





HOW ?!

- DDL Statement
- Primary Key + Foreign Key
- Special syntax to import CSV file



```
CREATE TABLE `Employee` (  
  `id` int PRIMARY KEY,  
  `first_name` varchar(20),  
  `middle_initial` varchar(20),  
  `last_name` varchar(20),  
  `gender` varchar(5),  
  `email` varchar(200),  
  `age` int,  
  `phone_no` character(12),  
  `date_of_birth` datetime  
);
```

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Data/hr_database/Employee.csv'  
INTO TABLE Employee  
FIELDS TERMINATED BY ','  
IGNORE 1 ROWS;
```