





DATA MANAGEMENT Assignment 3

Students:

- Andrea Potì 2008416
- Giulio D'Erasmo 1859130



The Dataset: Human Resource

HR: Analytics

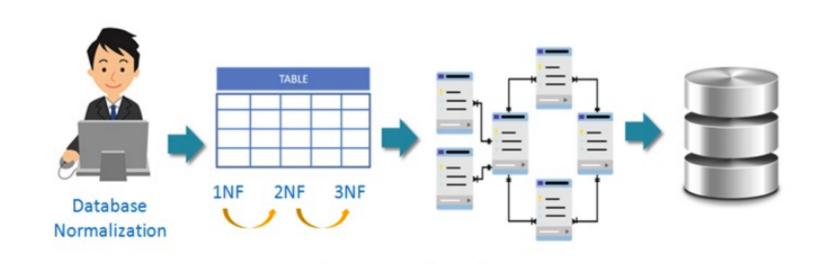
A٤	ge	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
Ec	ducation Field	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
Jo	bLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked
O	ver18	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StandardHours	StockOptionLevel
To	otalWorkingYears	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					

SQL

Avoid redundancy of data:
 NORMALIZATION



Fixed SCHEMA + Constrain

Address

id_empintcountryvarchar(50)cityvarchar(50)statecharacter(2)zipintssncharacter(12)regionvarchar(50)

Employee

Login

id_emp int
username varchar(50)
password varchar(50)

Education

id_empinteducation_nointeducational_fieldvarchar(20)

ExtraInfoEmployee

• •	
id_emp	int
fathers_name	varchar(50)
mothers_name	varchar(50)
mothers_last_name	varchar(50)
weight	int
marital_status	varchar(20)
number_company_worked	int
total_working_years	int
work_life_balance	int
distance_from_home	int
years_in_company	int

SatisfactionAndProductivity

hire_id int
job_satisfaction int
enviroment_satisfaction int
job_involment int
business_travel varchar(100)
performance_rating int

..

int first_name varchar(20) middle_initial varchar(20) last_name varchar(20) gender varchar(5) varchar(200) email age int phone_no character(12) date_of_birth datetime

Hire

id_emp int

hire_id int

department varchar(40)

job_role varchar(40)

jobLevel int

attrition varchar(3)

hire_date date

Salary

id_emp	int	
hire_id	int	
daily_rate	int	
hourly_rate	int	
monthly_rate	int	
monthly_income	int	
percent_salary_hike	int	

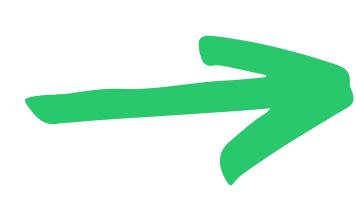
ExtraInfoHire

hire_id int
quarter_of_joining character(2)
half_of_joining varchar(5)
year_of_joining int
month_of_joining varchar(10)
day_of_joining int
day_name varchar(10)

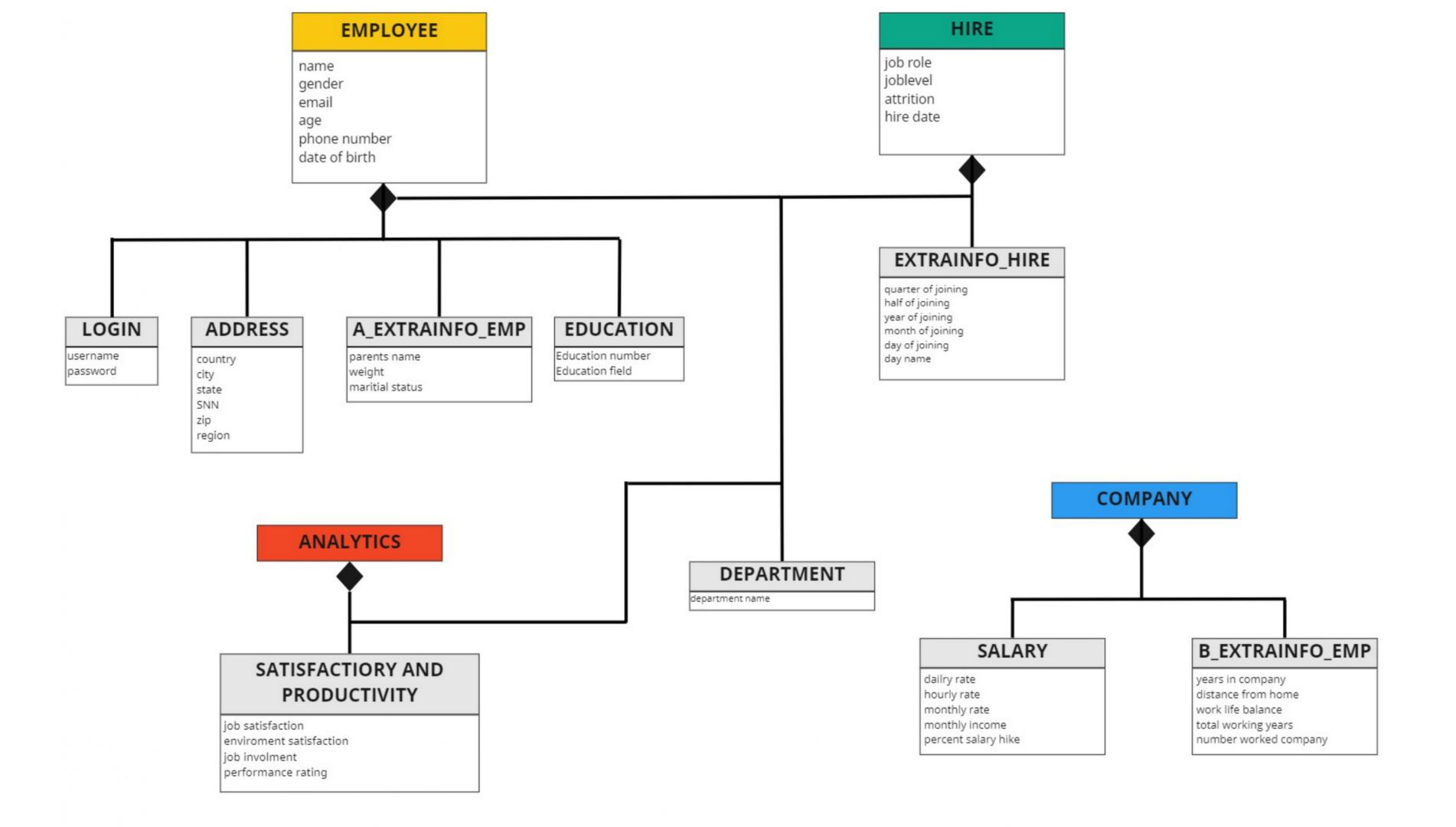
OLD SCHEMA

Going to NOSQL:

- Lots of Group By;
- Count operator;
- Analytics;
- Document base information



- Query based;
- Denormalization;
- No schema: all the data may be stored under the employee key ID



AGGREGATE DB: Document-based

- 1. Use PyMongo to create the dictionary corresponding to the Collections;
- 2. Import the Collections;
- 3. Adjust Data Types;
- 4. Create Index;
- 5. Query

QUERY

andrea

1

)

4 lookup

7 lookup

8 brutta

giulio

3

5

6

9

10

SQL

1 Table based

2 Better for multi-rows transactions

3 Rigid schema

4 Joins tipically required

MONGO-DB

1 Document based

Better for unstructured data like documents or JSON

3 Flexible schema

4 Joins tipically NOT required