



DATA MANAGEMENT

Assignment 3

Students:

- Andrea Potì 2008416
- Giulio D'Erasmus 1859130



SAPIENZA
UNIVERSITÀ DI ROMA

The Dataset: Human Resource

HR: Analytics

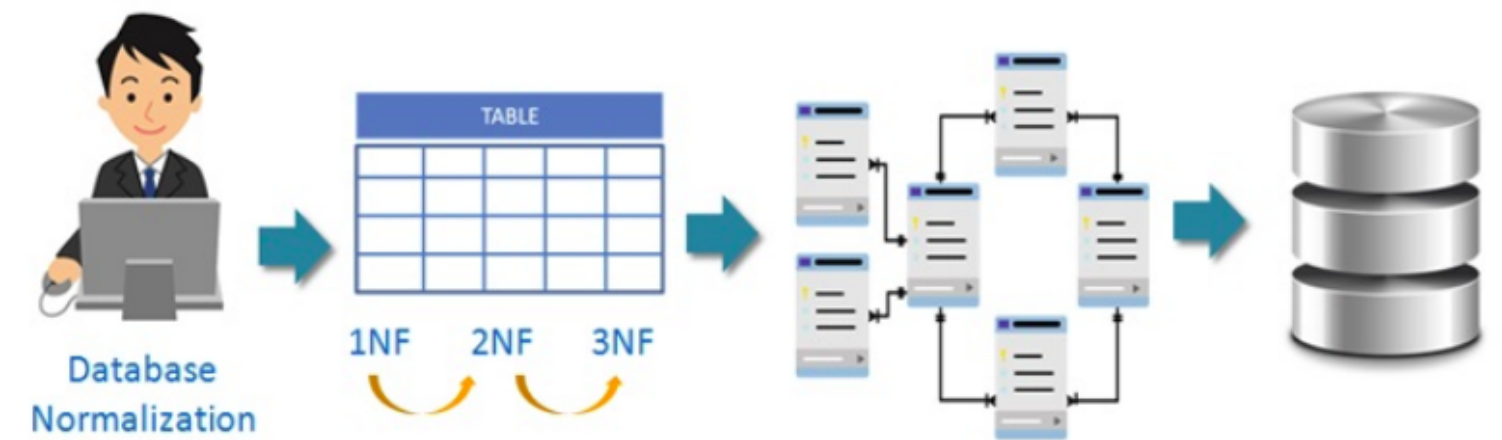
Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked
Over18	OverTime	PercentSalaryHike	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					

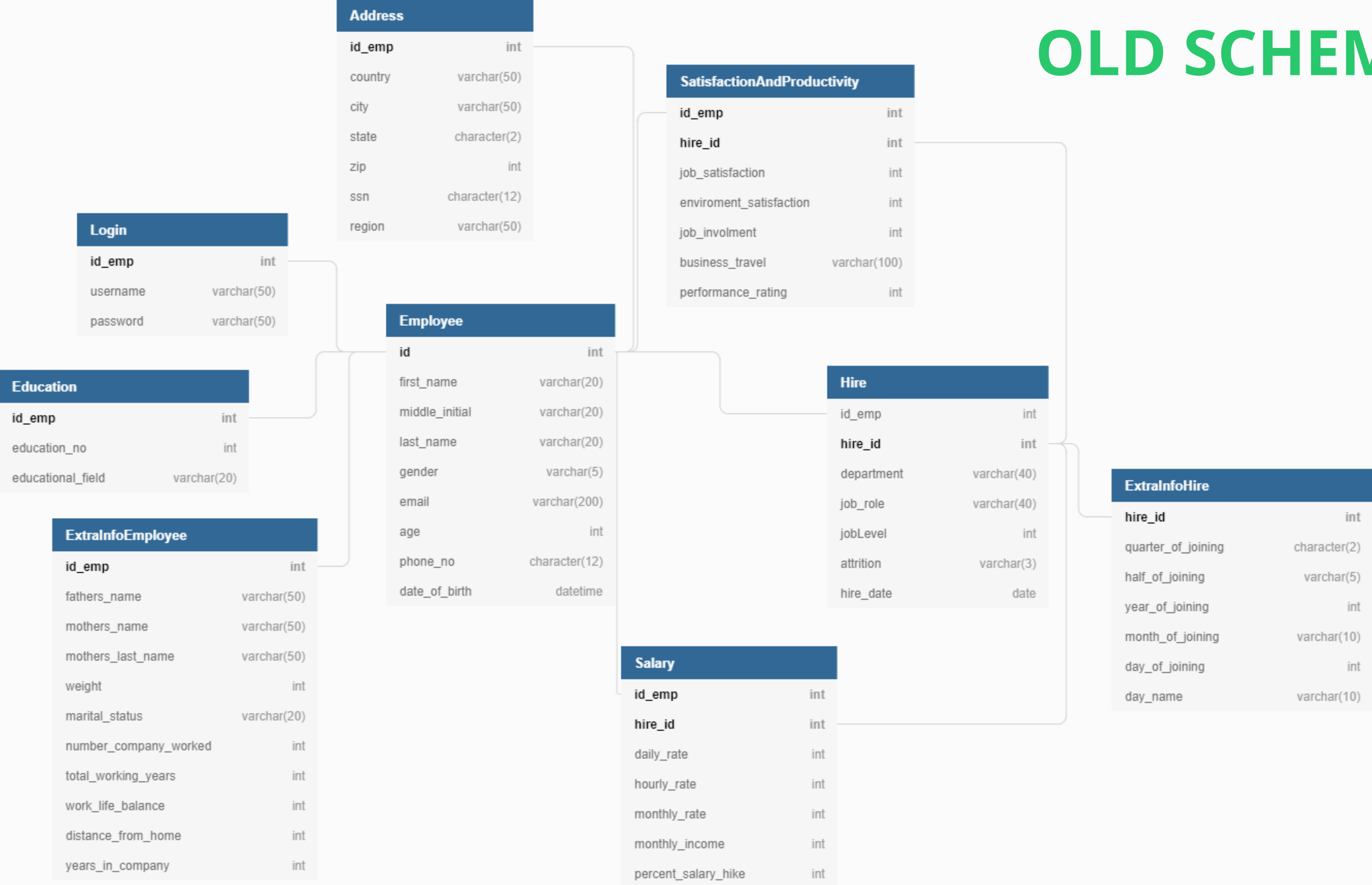
SQL

- Avoid redundancy of data:
NORMALIZATION



- Fixed **SCHEMA + Constrains**

OLD SCHEMA

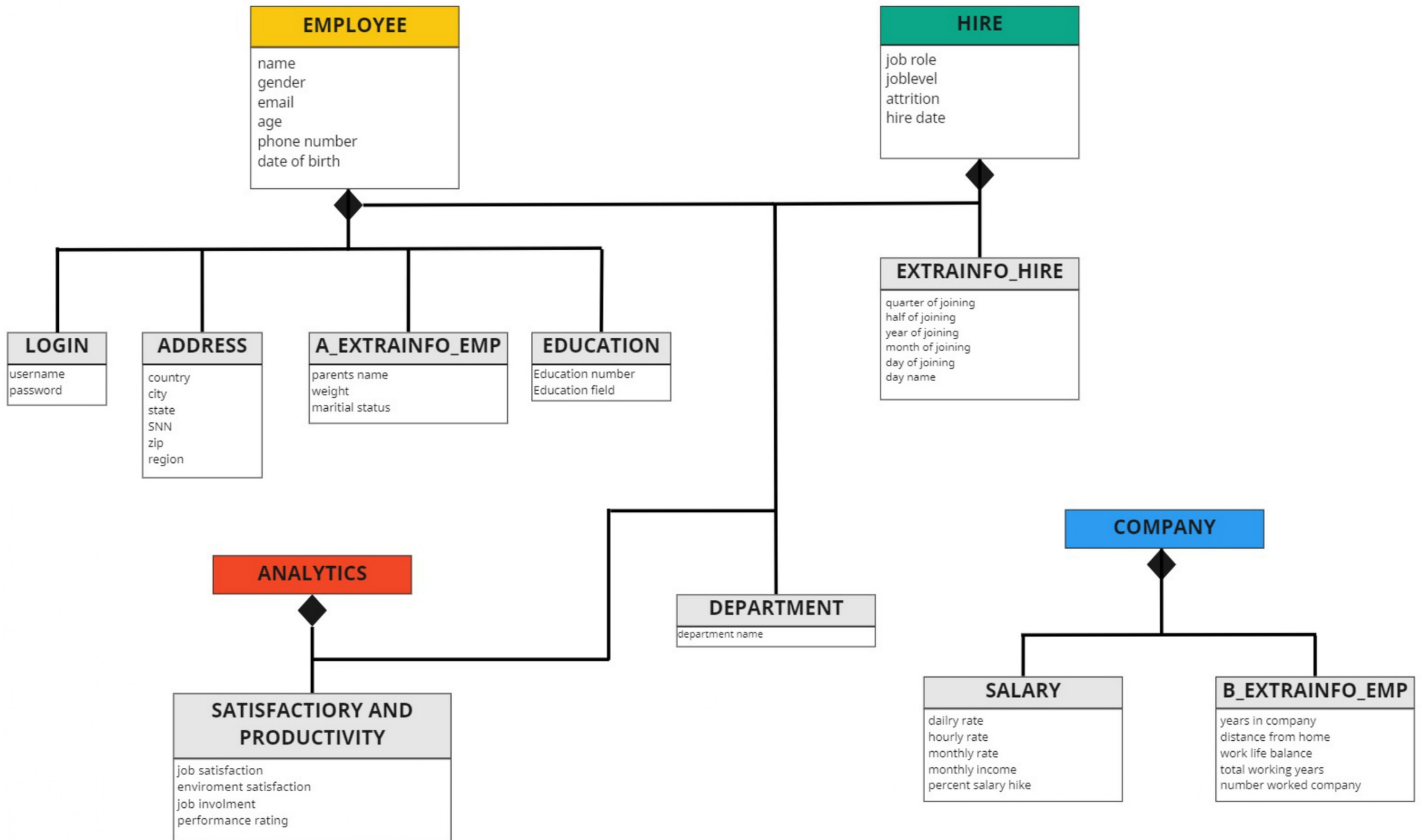


Going to NOSQL:

- 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
- 2
- 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 typically required

MONGO-DB

1

Document based

2

Better for
unstructured data like
documents or JSON

3

Flexible schema

4

Joins typically NOT required