

Prise en compte de données manquantes

T. Bouadi

Université de Rennes I (IUT Lannion)

Juin 2022

Valeurs manquantes : le NULL

- 1 **En théorie** : dans une variable relationnelle, tous les attributs ont une valeur
- 2 **En pratique** : certaines valeurs peuvent être inconnues ou manquantes

⇒ On dit qu'elles sont à **NULL**

Le NULL n'est pas une valeur spéciale, c'est une absence de valeur.

Trois valeurs de vérité

- ▶ La présence d'un NULL renvoie un résultat qui n'est ni TRUE ni FALSE mais UNKNOWN

Trois valeurs de vérité: vrai, faux et **inconnu**

- 1 vrai AND inconnu = inconnu
- 2 faux AND inconnu = faux
- 3 inconnu AND inconnu = inconnu
- 4 vrai OR inconnu = vrai
- 5 faux OR inconnu = inconnu
- 6 inconnu OR inconnu = inconnu
- 7 NOT inconnu = inconnu

Exemple

SCHÉMA : **EMPLOYEE**(EMPNO,ENOM,DEPNO,SAL)

SQL:

```
SELECT ENOM
FROM EMPLOYEE
WHERE SAL > 20000 OR SAL <= 20000
```

On ne trouve que les noms des employés avec un salaire connu. Pourquoi?

Comparaison avec des valeurs NULL

La valeur NULL est une valeur “spéciale” qui représente une *valeur (information) inconnue*.

- ① $A \theta B$ est inconnu (ni vrai, ni faux) si la valeur de A ou/et B est NULL (θ est l'un de $=, <, >, \neq, \dots$).
- ② $A \text{ op } B$ est NULL si la valeur de A ou/et B est NULL (op est l'un de $+, -, *, /$).

Exemple

SCHÉMA et INSTANCE :

FOURNISSEUR	FNOM	VILLE
	Toto	Paris
	Lulu	NULL
	Marco	Marseille

REQUÊTE: *Les Fournisseurs de Paris.*

SQL:

```
SELECT FNOM
FROM   FOURNISSEUR
WHERE  VILLE = 'Paris'
```

RÉPONSE : Toto

Exemple (suite)

REQUÊTE: *Fournisseurs dont la ville est inconnue.*

SQL:

```
SELECT FNOM  
FROM   FOURNISSEUR  
WHERE  VILLE = NULL
```

La réponse est vide. Pourquoi?

SQL:

```
SELECT FNOM  
FROM   FOURNISSEUR  
WHERE  VILLE IS NULL
```

RÉPONSE : Lulu

Valeur NULL vs Chaîne vide

Pour les données textuelles, on peut également retrouver des chaînes vides. La *distinction entre la signification d'une valeur NULL dans un champ et d'une chaîne vide est subtile* :

- ▶ La valeur **NULL** signifie que la valeur est **absente** (non saisie)
- ▶ La **chaîne vide**, quant à elle, correspond à une valeur qui **n'existe pas** (un·e étudiant·e n'ayant pas de deuxième prénom)