

# BASES DE DONNÉES AVANCÉES

## Clés et Formes Normales

### Travaux dirigés

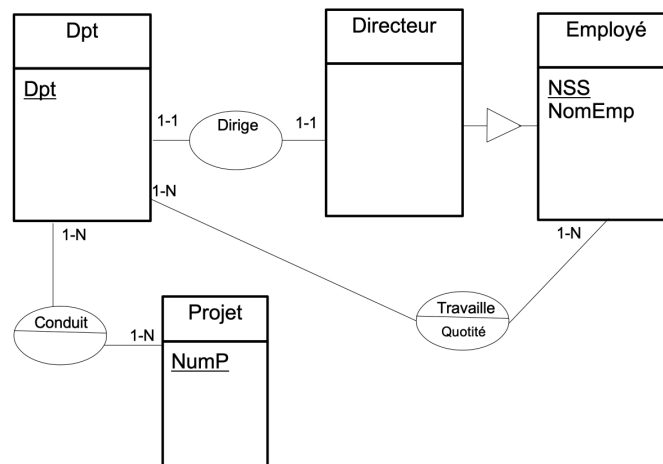
#### Exercice 1 : Tests de formes normales

Soit  $R = ABCDE$  un schéma de relation sur lequel on définit un ensemble de Dépendances Fonctionnelles et de Dépendances de Jointure  $\Sigma$ . Dans chacun des cas suivants, a) déterminez les clés minimales de  $(R, \Sigma)$ , b) donnez la meilleure forme normale dans laquelle se trouve  $R$  : 1FN, 3FN, FNBC, 4FN, 5FN et c) Proposez de façon intuitive ce qui vous semblerait être une "bonne" décomposition.

**Rappel** : toute clé minimale contient au moins les attributs qui n'apparaissent en partie droite d'aucune DF. Pour tester qu'un ensemble donné est bien une clé, on peut calculer sa fermeture.

1.  $\Sigma = \{ABC \rightarrow E\}$
2.  $\Sigma = \{A \rightarrow E; B \rightarrow D; D \rightarrow C; D \rightarrow A\}$
3.  $\Sigma = \{D \rightarrow A; A \rightarrow B; A \rightarrow C; A \rightarrow E\}$
4.  $\Sigma = \{A \rightarrow B; B \rightarrow A; C \rightarrow D; C \rightarrow E\}$
5.  $\Sigma = \{ABC \rightarrow D; ABC \rightarrow E; E \rightarrow A\}$
6.  $\Sigma = \{ABCD \rightarrow E; E \rightarrow ABCD\}$
7.  $\Sigma = \{ABCD \rightarrow E; E \rightarrow ABCD; \bowtie [ACE, BCDE]\}$

#### Exercice 2 : Entités / Associations et DF



Soit une application dont les données sont modélisées par le schéma ci-dessus. On suppose une traduction dans une relation unique :

$EMP\_DPT(Dpt, NSS\_Dir, NSS, NomEmp, Quotité, NumP)$

1. En vous appuyant sur le schéma E/A, faites l'inventaire des DF valides sur ce schéma.
2. Déduisez-en les clés de  $EMP\_DPT$

3. Dans quelle forme normale se trouve ce schéma ?
4. Dans quelle forme normale se trouve le schéma :

$EMP\_DPT(Dpt, NSS\_Dir, NSS, Quotité, NumP); EMP(NSS, NomEmp)$

5. Considérons la relation  $EMP\_DPT(Dpt, NSS, Quotité, NumP)$ . On sait qu'un département est associé à plusieurs employés.
  1. Peut-on dire que la DMV  $Dpt \rightarrow NSS$  est valide dans cette relation ? Pourquoi ?
  2. Qu'en est-il pour la même dépendance dans la relation  $EMP\_DPT(Dpt, NSS, NumP)$  ?
6. Donnez un schéma normalisé en appliquant l'algorithme de traduction ; vérifiez que dans le schéma obtenu, toutes les DF inventoriées sont bien respectées.

### Exercice 3 : Retrouver la sémantique d'une relation

Reprises	NUMFILM	TITRE	ANNEE	NUMORIGINAL	ANNEEORIGINAL	SIMILARITE
	5050	Dracula	1992	2123	1931	0,2
	2123	Dracula	1931	NULL	NULL	NULL
	3434	Une messe pour Dracula	1970	2123	1931	0,3
	1470	Dracula	2012	5050	1992	0,5
	5050	Dracula	1992	3434	1970	0,05
	1300	Les cicatrices de Dracula	1970	2123	1931	0,1

Soit la relation *Films* représentée dans la table ci-dessus, dont on ne possède pas le cahier des charges et la conception.

1. Cette relation vous paraît-elle normalisée ? Mettez en avance quelques redondances probablement liées à des DF.
2. Imaginez, à partir des noms d'attributs, la "signification" de la relation. Vérifiez ou écarter vos hypothèses en vous appuyant sur les DF et les Dépendances d'Inclusion qui seraient vérifiées ou contredites.
3. Dressez le schéma E/A qui pourrait correspondre à cette relation. Représentez la base de données "correcte" en remplissant les tables avec les bons tuples.

## Corrections

### Solution de l'exercice 1

1. A, B, C et D n'apparaissent dans aucune partie droite ; donc toute clé minimale contient ces attributs. Or, ABCD permet d'identifier E puisque  $ABC \rightarrow E$ , donc ABCD est la seule clé minimale.  
Dans la DF  $ABC \rightarrow E$ , la partie gauche n'est pas clé donc on n'est pas en FNBC. Puisque E n'appartient pas à une clé minimale, on n'est pas en 3FN. On retient donc uniquement la 1FN.
2. B n'est dans aucune partie droite donc appartient à toute clé minimale. Par ailleurs, on voit que B permet d'identifier D, puis par transitivité, C, A puis E. Donc B est la seule clé minimale.  
On voit que  $A \rightarrow E$ , or A n'est pas une clé, donc pas FNBC. Pour la même DF, E n'appartient pas à une clé minimale, donc on n'est pas en 3FN. On retient donc uniquement la 1FN.
3. D est dans toutes les clés minimales. D permet d'identifier A, B, C et E donc D est la seule clé minimale unique. Ainsi, la DF  $A \rightarrow B$  permet d'affirmer qu'on est seulement en 1FN.
4. C est dans toutes les clés minimales. C permet d'identifier D et E uniquement. La seule façon d'identifier A et B est d'ajouter A ou B à la clé ; on voit donc que AC et BC permettent d'identifier l'ensemble de la relation. Donc il y a deux clés minimales : AC et BC.  
On voit plusieurs DF qui ne satisfont pas la FNBC, car leur partie gauche n'est pas clé (par exemple  $A \rightarrow B$ ). Par ailleurs,  $C \rightarrow D$  ne satisfait pas la 3FN car D n'est pas dans une clé minimale. On est donc en 1FN.
5. B et C appartiennent à toutes les clés, mais BC n'est pas un identifiant de toute la relation, il faut étudier ses sur-ensembles. Si on considère ABC, on a bien une clé minimale. On se rend compte également que BCE permet d'identifier tous les attributs car  $E \rightarrow A$ , donc BCE permet d'identifier ABC. On a deux clés minimales : ABC et BCE.  
Seule la DF  $E \rightarrow A$  contredit la FNBC puisque E n'est pas clé. Comme A est un attribut d'une clé minimale, cette DF respecte en revanche la 3FN. Donc le schéma se trouve en 3FN.
6. On se rend compte facilement que les clés minimales sont E et ABCD.  
Le schéma est en FNBC ; comme il n'y a pas de dépendance de jointure, on peut affirmer qu'on est en 5FN.
7. Idem ; mais cette fois il y a une dépendance de jointure. Celle-ci est un cas de dépendance multivaluée qui peut s'écrire sous la forme  $CE \twoheadrightarrow A$ . Or, la DF  $CE \rightarrow A$  est bien valide ; donc la DJ ici est en réalité assurée par la DF. On est bien en 5FN.

### Solution de l'exercice 2

1. On lit sur le schéma :  $\{NSS \rightarrow NomEmp; NSS\_Dir \rightarrow Dpt; Dpt \rightarrow NSS\_Dir; Dpt, NSS \rightarrow Quotité\}$
2. On voit que les clés contiennent  $\{NSS, NumP\}$  ; les attributs NSS\_Dir et Dpt sont "équivalents". On arrive à la conclusion qu'il y a deux clés minimales :  $(NSS, NumP, NSS\_Dir)$  et  $(NSS, NumP, Dpt)$ .
3. Le schéma est en 1FN car la DF  $NSS \rightarrow NomEmp$  contredit la FNBC et la 3FN.
4. La relation EMP est en FNBC car NSS est clé. On voit que la relation EMP\_DPT est en première forme normale.
5. 1. Non, car Dpt ne correspond pas à un ensemble d'employés indépendamment de la quotité. Intuitivement, si deux tuples portent sur le même département, on ne peut pas intervertir les valeurs sur NSS ; on contredirait la DF  $Dpt, NSS \rightarrow Quotité$ .  
2. Ici il y a bien indépendance, pour un département donné, entre ses projets et ses employés donc la DMV est valide. On le voit sur le schéma E/A.  
**Remarque en conclusion** : La validité d'une DF est absolue, indépendamment de la relation où elle s'exprime. En revanche, la validité d'une DMV dépend du schéma de relation dans lequel on l'exprime, en raison de la notion d'indépendance.

6. La traduction donne (les clés étrangères sont omises) :

$EMP\_DPT(Dpt, NSS, Quotité); EMP(NSS, NomEmp); DPT(Dpt, NSS\_Dir)$

$DIR(NSS\_Dir); DEP\_PROJ(Dpt, NumP)$   
Chaque relation est bien en 5FN donc la BD est en 5FN.

### Solution de l'exercice 3

1. On voit que l'année est répétée pour un même numéro de film. On suspecte donc, par exemple, la DF  $NUMFILM \rightarrow ANNEE$  d'être valide (parmi d'autres!). Or  $NUMFILM$  n'est pas clé car il y a des doublons sur cet attribut, donc la relation n'est pas normalisée.
2. Grâce aux DF satisfaites et non satisfaites, on se rend compte que chaque film apparaît plusieurs fois (toujours avec le même titre et année,  $NUMFILM \rightarrow TITRE, ANNEE$  est satisfaite), et qu'il correspond à plusieurs originaux. On a une dépendance d'inclusion satisfaite entre  $NUMORIGINAL$  et  $NUMFILM$ , ce qui confirme que les originaux sont bien des films existants. Les originaux ont été repris par plusieurs films. Chaque original semble correspondre à une seule année ( $NUMORIGINAL \rightarrow ANNEEORIGINALE$ ). La similarité est une caractéristique d'un couple (film, original) : on peut dire que  $\{NUMFILM, NUMORIGINAL\}$  semble être la clé de cette relation; à la nuance près qu'il y a des valeurs NULL dans l'attribut  $NUMORIGINAL...$  (Ce qui est une conséquence parmi d'autres de la dénormalisation de cette relation.)
3. Schéma ci-dessous :

