UNIVERSITE BADJI MOKHTAR-ANNABA
FACULTÉ DES SCIENCES DE L'INGENIEUR
DÉPARTEMENT D'INFORMATIQUE

Normalisation

BDD 2LMD

Partie 1

Présenté par : Dr BELLEILI Habiba

Plan

- Définitions
- Redondances
  - Anomalies
- Dépendances Fonctionnelles
- Axiomes d’Amstrong
- DF élémentaire / DF augmentée
- Réécriture de DF
- Fermeture transitive
- Couverture minimale / Graphe minimum
- Clé d’une relation

Définitions

- La théorie de la normalisation est une théorie destinée à concevoir un **bon** schéma d'une base de données sans redondance d'information et sans risques d'anomalie de mise à jour.
- **Redondance d'information**: les informations sont *répétées à plusieurs endroits* de la base de données
- Elles constituent des sources de problèmes puisqu'elles sont l'une des Causes **d'incohérence** dans la BDD

Redondances

Livraison (Nofourn, adrF, Noprod, couleur, prixP, date, quantité)

NoFourn	adrF	Noprod	couleur	prixP	date	quantité
1233	**Annaba**	**P56**	**rouge**	147	12/09/19	5
1582	Skikda	P963	vert	245	16/10/19	2
1233	**Annaba**	P69	noir	159	30/11/19	3
1698	Alger	**P56**	**rouge**	258	15/12/19	2
1582	Skikda	P33	noir	168	10/01/20	5
1233	**Annaba**	**P56**	**rouge**	147	12/01/20	3

Anomalies de mise à jour

NoFourn	adrF	Noprod	couleur	prixP	date	quantité
1233	~~Sétif~~ Skikda	P56	rouge	147	12/09/19	5
1582	Annaba	P963	vert	245	16/10/19	2
1233	Sétif	P69	noir	159	30/11/19	3
1698	Alger	P56	rouge	258	15/12/19	2
1582	Skikda	P33	noir	168	10/01/20	5
1233	Sétif	P56	rouge	147	12/01/20	3

incohérence

Redondances → mise à jour simple → INCOHÉRENCE

Redondances → mise à jour COMPLEXE

Anomalies d'insertion

- A chaque nouvelle livraison je dois insérer le numéro de fournisseur et son adresse
- Si je me trompe dans l'adresse je vais avoir une incohérence

1233	Annaba	P56	rouge	147	12/03/20	3

Redondances

- La relation qui présente des redondances est incorrecte

- Il faut la **décomposer**

- On parle alors de **normalisation**
- La normalisation d'une relation est un processus de décomposition d'une relation présentant des mise à jour **complexes en plusieurs relations à mises à jour simples.**

Exemple

Fournisseur (N°fourn, adrF)

Produit (N°prod, couleur)

NoFourn	adrF	Noprod	couleur	prixP	date	quantité
1233	Annaba	P56	rouge	147	12/09/19	5
1582	Skikda	P963	vert	245	16/10/19	2
1233	Annaba	P69	noir	159	30/11/19	3
1698	Alger	P56	rouge	258	15/12/19	2
1582	Skikda	P33	noir	168	10/01/20	5
1233	Annaba	P56	rouge	147	12/01/20	3

NoFourn	adrF
1233	Annaba
1582	Skikda
1698	Alger

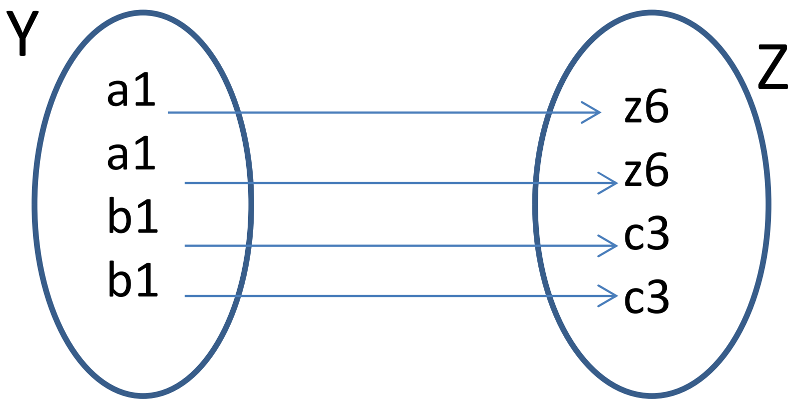
Noprod	couleur
P56	rouge
P963	vert
P69	noir
P33	rouge

Livraison (N°fourn, N°prod, Prix, date, qté)

NoFourn	Noprod	prixP	date	qté
1233	P56	147	12/09/19	5
1582	P963	245	16/10/19	2
1233	P69	159	30/11/19	3
1698	P56	258	15/12/19	2
1582	P33	168	10/01/20	5
1233	P56	147	12/01/20	3

Notion de **D**épendances **F**onctionnnelles (DF)

- Définition: Les DF permettent d'établir des **liens sémantiques** entre attributs ou groupe d'attributs
- Soit R une relation de schéma R(X,Y,Z,…)
- il existe une **"DF", de Y vers Z,** notée Y→Z, si :
- Etant donné *deux tuples* quelconques de R,
- s'ils ont même valeur pour Y, alors ils ont **nécessairement** même valeur pour Z.



On appelle Y **source** de la DF, et Z **cible** de la DF

Y → Z

Exemple

A	B	C	D	E
a1	b2	c2	d3	e2
a1	b2	c2	d1	e4
a2	b3	c2	d1	e5
a2	b4	c5	d1	e5

$A \rightarrow B$ n’est pas une DF

$B \rightarrow A$ est une DF

$D \rightarrow E$ n’est pas est une DF

$E \rightarrow D$ est une DF

Propriétés des DF

Axiomes d'Armstrong

Réflexivité: Soient X et Y des attributs :

$X \rightarrow X$ $\quad$ $XY \rightarrow X$ $\quad$ $XY \rightarrow XY$ $\quad$ **triviales**

$Y \rightarrow Y$ $\quad$ $YX \rightarrow Y$

Augmentation: Soient X, Y et Z des attributs :

$$X \rightarrow Y \Longrightarrow \begin{cases} XZ \rightarrow YZ \\ XZ \rightarrow Y \end{cases}$$

Transitivité: Soient X, Y et Z des attributs :

$$\left. \begin{array}{l} X \rightarrow Y \text{ et} \\ Y \rightarrow Z \end{array} \right\} \Longrightarrow X \rightarrow Z$$

Axiomes d'Amstrong

- **Pseudo-transitivité**: Soient, W, X, Y et Z des attributs

  $X \rightarrow Y$ et $WY \rightarrow Z \Rightarrow WX \rightarrow Z$

  Démonstration:

  $X \rightarrow Y \Rightarrow WX \rightarrow WY$ (réflexivité)

  $WX \rightarrow WY$ et $WY \rightarrow Z \Rightarrow WX \rightarrow Z$ (transitivité)

- **Union**: Soient X, Y et Z des attributs

  $X \rightarrow Y$ et $X \rightarrow Z \Rightarrow X \rightarrow YZ$

  Démonstration

  transitivité

  Réflexivité triviale

  Augmentation $X \rightarrow Y$

  $(X \rightarrow Y$ et $X \rightarrow Z \Rightarrow X \rightarrow XX$ et $XX \rightarrow XY \Longrightarrow X \rightarrow XY$

  $X \rightarrow XY$ et $YX \rightarrow YZ \Rightarrow X \rightarrow YZ)$

  Augmentation $X \rightarrow Z$

  transitivité

DF élémentaire

- une DF, X -> B est **élémentaire si B est un attribut** unique, et si X est un ensemble minimum d'attributs (ou un attribut unique)
  - NoFourn,Noprod,date → quantité est élémentaire
- DF non élémentaires:
  - Côté source :
    - J'ai A→B alors AX→B est non élémentaire (augmentée)
    - la cible est inclus dans la source : AB→A $A \subset AB$. Elle est dite triviale
  - Côté cible : la cible est un groupe d'attributs AB→C,B C,B est un groupe d'attributs.

DF augmentée

- Une DF **non élémentaire** est dite Augmentée : si X→Y alors quelque soit A A,X→ Y est une DF Augmentée
  - NoFourn→adrF   est élémentaire
  - NoFour,Noprod→adrF   est augmentée

Réécriture de DF

- On peut toujours réécrire un ensemble de DF en un ensemble de DFE (DF élémentaires):
  - en supprimant les DF triviales obtenues par **réflexivité,**
  - en décomposant les DF à **partie droite non atomique** en plusieurs DFE
- AB→A n'est pas considérée car c'est une DF triviale obtenu par réflexivité.
- A→B,C sera réécrite: A→B et A→C
- AB→CB est décomposée en
  - AB→C et
  - ~~AB→B~~

triviale

Fermeture Transitive

- On appelle fermeture transitive F+ d'un ensemble F de DFE, l'ensemble de toutes les DFE qui peuvent être **composées par transitivité** ou **pseudo transitivité** à partir des DFE de F
  - F = {A→B, B→C, B→D, A→E}.
  - La fermeture transitive de F est F+ = { A→B, B→C, B→D, A→E, A→C, A→D }
- La fermeture transitive permet de retrouver toutes les DFE

Couverture minimale des DFE

- La couverture minimale (CM)d'un ensemble de DFE (notée DFE*) est un sous-ensemble minimum des DFE permettant de générer toutes les autres DFE.
- Tout ensemble de DFE admet au moins une CM
- Un ensemble de DFE peut avoir plusieurs CM

$F=\{A \rightarrow B,\ A \rightarrow C, B \rightarrow C, C \rightarrow B\}$

$\left.\begin{array}{l} A \rightarrow B \\ B \rightarrow C \end{array}\right\} \cancel{A \rightarrow C}$ DFE*=$\{A \rightarrow B,\ \cancel{A \rightarrow C},\ B \rightarrow C,\ C \rightarrow B\}$

2 Couvertures Minimales

$\left.\begin{array}{l} A \rightarrow C \\ C \rightarrow B \end{array}\right\} \cancel{A \rightarrow B}$ DFE*=$\{\cancel{A \rightarrow B},\ A \rightarrow C,\ B \rightarrow C,\ C \rightarrow B\}$

Couverture minimale: Algorithme

- **Entrée:** *F* un ensemble de dépendances fonctionnelles
- **Sortie:** G une couverture minimale de *F*
- **Début**

1.G := *F*

2.**Décomposer**: Pour chaque DF ∈ G , **appliquer la règle de décomposition** (axiome d'Armstrong)

X-->ABC   sera décomposé en X-->A ; X-->B; X-->C

3. Déterminer les **DFs élémentaires en supprimant les DF augmentées**: Supprimer les attributs en <u>surnombre</u> à gauche :

Pour tout X --> Y, s'il existe dans G un Z⊆X tel que Z-->Y alors remplacer X-->Y par Z-->Y

4. **Supprimer les DF déduites** :

Une DF X-->A est déduite si elle peut être retrouvée par transitivité ou pseudo transitivité

si X-->Z et Z-->A alors (par transitivité) X-->A

si X-->Y et Y,Z-->A alors (par pseudo transitivité) X,Z --> A   voir diapo 12

Fin

Graphe minimum des DF

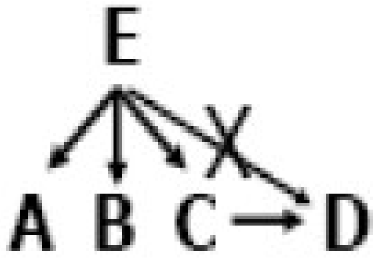
- On appelle **graphe minimum des DF de la relation, tout ensemble de DF élémentaires non déduites,**
  - **DF élémentaires → Pas de DF augmentée**
  - **DF non déduite → par transitivité**

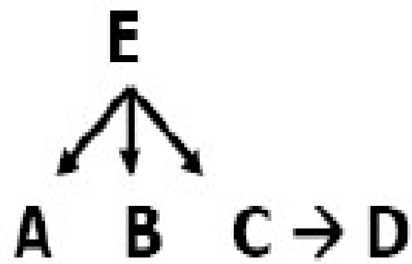
- les DF **augmentées et déduites doivent être supprimer du graphe des DF pour** obtenir un graphe minimum des DF.

Le graphe minimum des DF sert essentiellement à définir des relations normalisées.

DF déduite ?

- Une méthode pour savoir si une DF, X→ Y, est déduite des autres DF est la suivante:
  - établir un graphe de toutes les DF, (non minimum)
  - supprimer la DF X→Y du graphe,
  - parcourir tous les chemins possibles partant de X et suivant les DF. La DF, X→Y, est déduite si un (ou plusieurs) de ces chemins atteint Y.

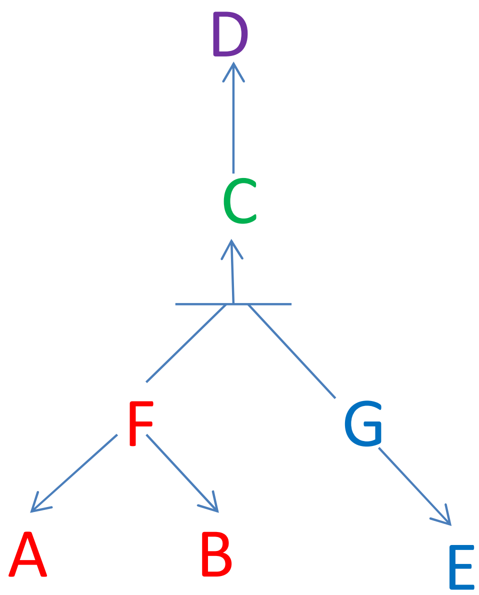


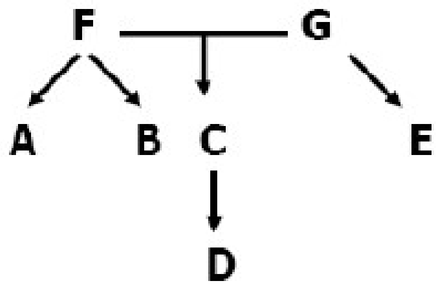


Exemple

- R(A,B,C,D,E,F,G)

DF={F→A,
F→B,
G→E,
F,G→C,
C→D}





Clé d’une relation

- La clé d’une relation peut être cherchée à partir d’un ensemble initial de DF (Couverture Minimale ou quelconque)
- Les méthodes pour trouver la clé d’une relation:
  - L’ Algorithme de la fermeture transitive sur **les attributs** via l’ensemble de DF
  - Ou en appliquant les règles d’Armstrong sur l’ensemble des DF
  - Ou intuitivement
  - à partir de la **superclé**

Clé: Algorithme par fermeture transitive d'un attribut

- ***Données:*** *F un ensemble de DF et X un ensemble d'attributs*
- ***Résultat:*** *X^+ fermeture transitive de X*
- ***Algorithme de saturation:***

1. *Initialiser $(X)^+$ à X,*
2. *Trouver une DF $\in$ F possédant en partie gauche des attributs inclus dans $(X)^+$,*
3. *Ajouter dans $(X)^+$ les attributs placés en partie droite de la DF*
4. *Répéter les étapes 2) et 3) jusqu'à ce que $(X)^+$ n'évolue plus.*

Clé par fermeture transitive sur les attributs

- R(A,B,C,D,E,F)
- F={AB→C, C→A,BC→D, ACD→B, BE→C, CE→BD, CE→FA, D→EF}
- Clé (s) de la relation R

$(C)^+$={C} // initialisation

- $1^{ière}$ itération $(C)^+$ ={C,A},
- $2^{ième}$ itération **$(C)^+$= {CA}** reste inchangé je m'arrête // C^+ ={C,A}

(D)+ = {D}

- $1^{ière}$ itération (D)+ ={D,E,F} (D→EF) ,
- $2^{ième}$ itération **(D)+ = {D,E,F}** reste inchangé je m'arrête

(AB)+ = {A,B}

- $1^{ière}$ itération (AB)+ ={A,B,C} (AB→C), (AB)+ ={A,B,C,D}(BC→D), (AB)+ ={A,B,C,D,E,F} (D→EF)
- $2^{ième}$ itération (AB)+= ={A,B,C,D,E,F} reste inchangé ARRET //de plus tous les attributs sont obtenus à partir de AB donc **(AB) une clé candidate**

(BC)+ ={B,C}

- $1^{ière}$ itération(B,C)+={BCAD} (C→A,BC→D) (BC)+={B,C,A,D,E,F} (D→EF)
- $2^{ième}$ itération (BC))+={B,C,A,D,E,F} reste inchangé ARRET // **(BC) clé candidate** (génère tous les attributs)

(BE)+ = {B,E}

$1^{ière}$ itération (BE)+={B,E,C} (BE→C), (BE)+={B,E,C,A,D,F},

$2^{ième}$ itération (BE)+={B,E,C,A,D,F} reste inchangé ARRET // **(BE) clé candidate**

(CE)+ = {C,E}

- $1^{ière}$ itération (CE)+={C,E,B,D,A,F},
- 2itération (CE)+={C,E,B,D,A,F}, reste inchangé ARRET **// (CE) clé candidate**

Superclé

- On Appelle une **superclé** d'une relation R une clé contenant:
  - Tous les attributs de la relation
  - Ou, pour optimiser, c'est l'union de toutes les parties gauches des DF

  - Exemples:
  - R(A,B,C,D,E) F={A→BD, CD→E, B→E}
  - Superclé= (ABCDE) ou (ABCD)

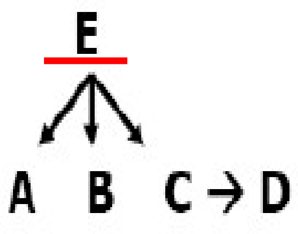
Clé par réduction de la superclé

- On peut obtenir la clé d'une relation par réduction de la superclé
  - Exemple:R(A,B,C,D,E) F={A→B, CD→E, B→E}
  - Superclé= (ABCDE)
  - ABCDE $\xrightarrow{A \to B}$ ACDE $\xrightarrow{CD \to E}$ ACD (2 étapes)
  - ACD est une clé

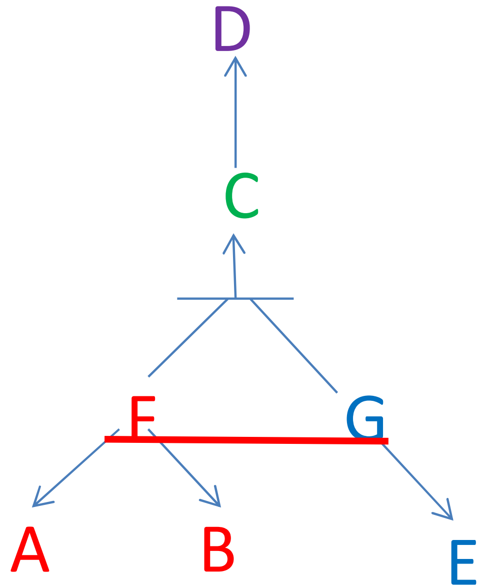
Une meilleure superclé est l'union des parties gauches des DF = ACD~~B~~ $\xrightarrow[A \to B]{}$ ACD (1 seule étape)

Clés d’une relation: a partir GM

- Les Clés peuvent aussi être cherchées à partir du graphe minimum des DF,
- Les Clés correspondent à l’ensemble minimum d’attributs qui nous permettent, en suivant, les DF d’atteindre tous les autres attributs.



La Clé est E



La Clé est (F+G)

Clés Candidates et clé primaires

- Si une relation comporte plusieurs clés, chacune est dite clé candidate
- On choisit une en particulier pour être la clé primaire.
- Toutes les clés candidates sont des clés, pas seulement la clé primaire.
- *Les clés candidates se déterminent mutuellement*
  - *Clé1 → Clé2 et aussi Clé2 → Clé1 (Clé1 ⟷ Clé2)*
- **Si** une relation R n'admet aucune clé K (sous ensemble des attributs A1..An de R)
- **alors** la clé K=A1..An est composée de *tous les attributs de R.*

Données: F un ensemble de DF et X un ensemble d'attributs
Résultat: X^+ fermeture transitive de X
Algorithme de saturation:
1. Initialiser $(X)^+$ à X,
2. Trouver une DF $\in$ F possédant en partie gauche des attributs inclus dans $(X)^+$,
3. Ajouter dans $(X)^+$ les attributs placés en partie droite de la DF
4. Répéter les étapes 2) et 3) jusqu'à ce que $(X)^+$ n'évolue plus.

Soit la relation R(A,B,C,D,E,F)
{AC $\rightarrow$ D, B $\rightarrow$AF, C$\rightarrow$ BE, F $\rightarrow$EC}
1) Donner les clés candidates par fermeture transitive sur les attributs.
2) donner les couvertures minimales

1) Clés candidates:

A+ = {A}

A+ ={A} 1ière itération A+ reste inchangé on arrête A n'est pas une clé candidate

B+ ={B} 1ière itération B+={B,A,F} (B→AF), 2ième itération B+={B,A,F,**E,C**}(F→EC), 3ième itération B+={B,A,F,E,C,**D**} (AC→D), 4ième itération B+ reste inchangé (on arrête)

TOUS les attributs sont générés donc B est une clé candidate

C+ ={C}, 1ière itération C+={C,B,E}(C→BE), 2ième itération C+={C,B,E,**A,F**}(B→AF), 3 ième itération C+={C,B,E,A,F,**D**}(AC→D), 4ième itération C+ reste inchangé donc on arrête de plus

C+ génère tous les attributs donc C est une clé candidate.

D+={D} 1ière itération D+={D} reste inchangé on arrête donc D n'est pas une clé candidate

E+={E} 1ière itération E+ reste inchangé arrêt donc E n'est pas une clé candidate

F+={F}, 1ière itération F+={F,**E,C**}, 2itération F+ ={F,E,C,**B** }, 3ième itération F+={F,E,C,B,**A** }, 4ième itération F+={F,E,C,B,A,**D**}, 5ième itération F+={F,E,C,B,A,D} reste inchangé on arrête.

F génère tous les attributs donc F est une clé candidate.

Les clés candidates sont B,C et F ⟹

B→ A,B,C,D,E,F
C→ A,B,C,D,E,F
F → A,B,C,D,E,F

2) couverture minimale ? {AC → D, B →**AF**, C→ **BE**, F →**EC**}

a) décomposition:

{ (1) AC--> D, (2)B-->A , (3) B-->F, (4)C-->B, (5)C--> E, (6) F-->E, (7) F-->C}

b) élimination des DF non élémentaires

seule la DF (1) a en partie gauche **2 attributs**, elle sera donc vérifiée si elle n'est pas élémentaire.

Les autres DF ont toutes un seul attribut (pas de surnombre) en partie gauche et donc ne peuvent pas être augmentées.

vérification de la DF (1) AC --> D

on calcule les fermetures transitives de A puis de C si la fermeture de A ou de C contient D alors AC --> D est augmentée.

A+={A}, voir première question

C+ ={A,B,C,D,E,F} car C est une clé candidate (1ière question) donc D $\in$ à C+ donc C-->D et C $\subset$ AC donc

AC--> D sera supprimée et remplacé par C-->D .

nouvel ensemble de DF { (1) C--> D, (2)B-->A , (3) B-->F, (4)C-->B, (5)C--> E, (6) F-->E, (7) F-->C} nouvel ensemble de DF { (1) C--> D, (2)B-->A , (3) B-->F, (4)C-->B, (5)C--> E, (6) F-->E, (7) F-->C}

c) supprimer les DF déduites

F-->C et C-->E donc F-->E par transitivité donc F--> E (6) est déduite (à supprimer)

nouvel ensemble de DF

{ (1) C--> D, (2)B-->A , (3) B-->F, (4)C-->B, (5)C--> E, ~~(6) F-->E~~,(7) F-->C} donc

1ière CM est { C--> D, B-->A , B-->F, C-->B, C--> E, F-->C}

On peut aussi trouver une autre couverture minimale en faisant deux transitivités:

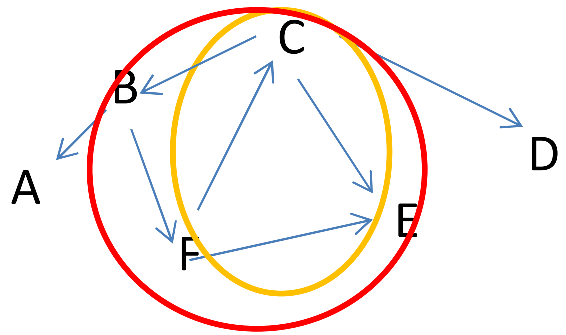
C-->B et B-->F par transitivité C-->F

C-->F et F--> E par transitivité C-->E donc C-->E est déduite (à supprimer)

nouvel ensemble de DF

{ (1) C--> D, (2)B-->A , (3) B-->F, (4)C-->B, ~~(5)C--> E~~, (6) F-->E,(7) F-->C} donc

2ième CM { C--> D, B-->A , B-->F, C-->B, F-->E,F-->C}.



Soit la relation

R(A,B,C,D,E,G,H}

F={

(1) $G \rightarrow A$,

(2) $AB \rightarrow C$,

(3) $B \rightarrow D$,

(4) $CD \rightarrow E$,

(5) $CE \rightarrow GH$}

ABCDEG --> ~~**A**~~BCDEG (car G --> A) ---> BCDEG (CE-->G)--> BCDE (CD-->E) --> BC D (B-->D) BC clé candidate

ABCDEG --> ABCDEG (CD-->E) ---> ABCDG (B-->D) ---> ABCG --> ABCG (B-->C) --> ABG (G-->A) BG clé candidate

ABCDEG --> ABCDGE (CE--> G) ---> ABCDE (CD -->E) --> ABCD (B-->D) ----> ABC (AB-->C) (AB) clé candidate

Fin Normalisation
Partie 1