

# 1. MOTIVATION ET INTRODUCTION

Version 3 - Septembre 2018

Support du chapitre 1, *Motivation et introduction*  
de l'ouvrage *Bases de données*, J-L Hainaut, Dunod 2018.

# 1. MOTIVATION ET INTRODUCTION

## Contenu

**1.1 Une première approche des données**

**1.2 Utiliser une base de données**

**1.3 Premières conclusions**

**1.4 Les systèmes de gestion de bases de données**

**1.5 Les défis des bases de données d'aujourd'hui**

**1.6 Un peu d'histoire**

**1.7 Et ensuite ?**

## 1.1 Une première approche des données

- Nous baignons dans un tourbillon d'informations
- Nous recevons des informations de notre environnement
- Nous transmettons des informations à notre environnement
- Les informations prennent la forme de données (*signes mémorisés et véhiculés sur un support matériel ou immatériel*)
- *Question* : comment stocker et traiter ces données ?

**Réponse** : dans une base de données

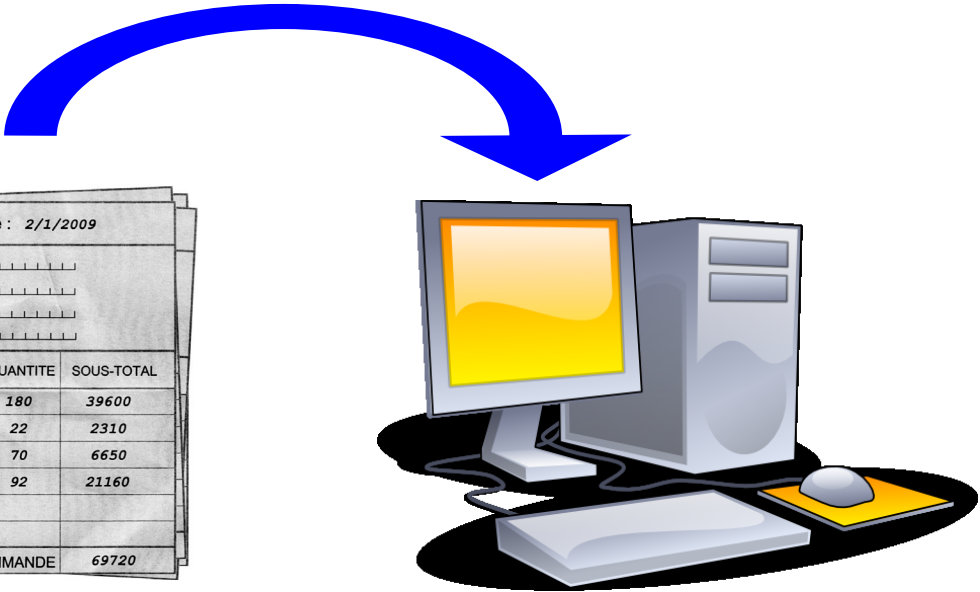
# 1.1 Une première approche des données

Les données sont omniprésentes autour de nous ...

Commande N° : 30188		Date : 2/1/2009		
Numéro client	B512			
Nom	GILLET			
Adresse	14, r. de l'Eté			
Localité	Toulouse			
N° PRODUIT	LIBELLE PRODUIT	PRIX	QUANTITE	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160
TOTAL COMMANDE				69720

## 1.1 Une première approche des données

Comment les ranger dans un ordinateur ?



Commande N° : 30188		Date : 2/1/2009		
Numéro client	B512			
Nom	GILLET			
Adresse	14, r. de l'Eté			
Localité	Toulouse			
N° PRODUIT	LIBELLE PRODUIT	PRIX	QUANTITE	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160
TOTAL COMMANDE				69720



## 1.1 Une première approche des données

### Comment les ranger dans un ordinateur ?

Commande N° : 30188		Date : 2/1/2009		
Numéro client : B512				
Nom : GILLET				
Adresse : 14, r. de l'Eté				
Localité : Toulouse				
N° PRODUIT	LIBELLE PRODUIT	PRIX	QUANTITE	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160
TOTAL COMMANDE				69720



- scanner les bons de commande [1/10]
- copier les données dans un tableau Word [2/10]
- copier les données dans une feuille Excel [4/10]
- copier les données dans une base de données [10/10]

# 1.1 Une première approche des données

En regardant d'un peu plus près ...

Commande N° : 30188		Date : 2/1/2009		
Numéro client		B512		
Nom		GILLET		
Adresse		14, r. de l'Eté		
Localité		Toulouse		
N° PRODUIT	LIBELLE PRODUIT	PRIX	QUANTITE	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160
TOTAL COMMANDE				69720

données du client

données de la commande

données d'un détail

# 1.1 Une première approche des données

Reportons ces données dans des tableaux :

Commande N° : 30188		Date : 2/1/2009		
Numéro client B512				
Nom GILLET				
Adresse 14, r. de l'Eté				
Localité Toulouse				
N° PRODUIT	LIBELLE PRODUIT	PRIX	QUANTITE	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160
TOTAL COMMANDE				69720

données du client

données d'un détail

données de la commande

données de la commande

NCOM	DATECOM	TOTAL-COMMANDE
30188	2/1/2009	69720

données du client

NCLI	NOM	ADRESSE	LOCALITE
B512	GILLET	14, r. de l'Eté	Toulouse

données des détails

NPRO	LIBELLE	PRIX	QCOM	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160



# 1.1 Une première approche des données

## Pas entièrement convaincant (1)

données de la commande

NCOM	DATECOM	TOTAL-COMMANDE
30188	2/1/2009	69720

données du client

NCLI	NOM	ADRESSE	LOCALITE
B512	GILLET	14, r. de l'Eté	Toulouse

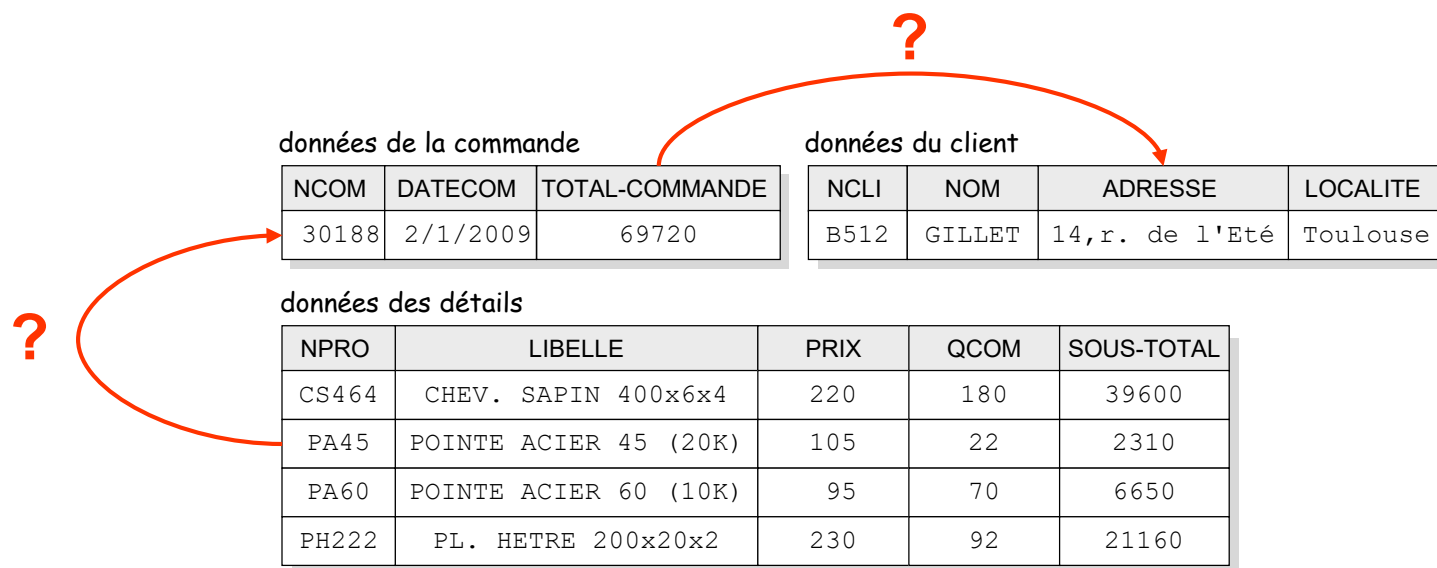
données des détails

NPRO	LIBELLE	PRIX	QCOM	SOUS-TOTAL
CS464	CHEV. SAPIN 400x6x4	220	180	39600
PA45	POINTE ACIER 45 (20K)	105	22	2310
PA60	POINTE ACIER 60 (10K)	95	70	6650
PH222	PL. HETRE 200x20x2	230	92	21160

Les données **TOTAL-COMMANDE** et **SOUS-TOTAL** sont calculées : inutile de les conserver, on pourra les recalculer en cas de besoin

# 1.1 Une première approche des données

## Pas entièrement convaincant (2)



**Il est impossible de reconstituer le bon de commande d'origine : quel est le client de la commande, quelle est la commande d'un détail ?**

## 1.1 Une première approche des données

### Données sans redondances (\*) et avec références :

données de la commande

NCOM	NCLI	DATECOM
30188	B512	2/1/2009

données du client

NCLI	NOM	ADRESSE	LOCALITE
B512	GILLET	14, r. de l'Eté	Toulouse

données des détails

NCOM	NPRO	QCOM	LIBELLE	PRIX
30188	CS464	180	CHEV. SAPIN 400x6x4	220
30188	PA45	22	POINTE ACIER 45 (20K)	105
30188	PA60	70	POINTE ACIER 60 (10K)	95
30188	PH222	92	PL. HETRE 200x20x2	230

(\*) presque !

# 1.1 Une première approche des données

Enregistrons les données de 2 bons de commandes :

données des commandes

NCOM	NCLI	DATECOM
30188	B512	2/1/2009
30179	C400	22/12/2008

données des clients

NCLI	NOM	ADRESSE	LOCALITE
B512	GILLET	14,r. de l'Eté	Toulouse
C400	FERARD	63,r. du Tertre	Poitiers

données des détails

NCOM	NPRO	QCOM	LIBELLE	PRIX
30188	CS464	180	CHEV. SAPIN 400x6x4	220
30188	PA45	22	POINTE ACIER 45 (20K)	105
→ 30188	PA60	70	POINTE ACIER 60 (10K)	95
30188	PH222	92	PL. HETRE 200x20x2	230
30179	CS262	60	CHEV. SAPIN 200x6x2	75
→ 30179	PA60	20	POINTE ACIER 60 (10K)	95

## Observation

si plusieurs détails mentionnent le même produit, ses caractéristiques sont répétées autant de fois : on isole les données des produits dans un tableau spécifique

# 1.1 Une première approche des données

## Distribution optimale des données des bons de commande

données des commandes

NCOM	NCLI	DATECOM
30188	B512	2/1/2009
30179	C400	22/12/2008

données des clients

NCLI	NOM	ADRESSE	LOCALITE
B512	GILLET	14,r. de l'Eté	Toulouse
C400	FERARD	63,r. du Tertre	Poitiers

données des détails

NCOM	NPRO	QCOM
30188	CS464	180
30188	PA45	22
30188	PA60	70
30188	PH222	92
30179	CS262	60
30179	PA60	20

données des produits

NPRO	LIBELLE	PRIX
CS464	CHEV. SAPIN 400x6x4	220
PA45	POINTE ACIER 45 (20K)	105
PA60	POINTE ACIER 60 (10K)	95
PH222	PL. HETRE 200x20x2	230
CS262	CHEV. SAPIN 200x6x2	75

## 1.1 Une première approche des données

### Ajoutons d'autres données : *notre première base de données*

COMMANDE		
NCOM	NCLI	DATECOM
30178	K111	22/12/2008
30179	C400	22/12/2008
30182	S127	23/12/2008
30184	C400	23/12/2008
30185	F011	2/01/2009
30186	C400	2/01/2009
30188	B512	2/01/2009

DETAIL		
NCOM	NPRO	QCOM
30178	CS464	25
30179	CS262	60
30179	PA60	20
30182	PA60	30
30184	CS464	120
30184	PA45	20
30185	CS464	260
30185	PA60	15
30185	PS222	600
30186	PA45	3
30188	CS464	180
30188	PA45	22
30188	PA60	70
30188	PH222	92

CLIENT					
NCLI	NOM	ADRESSE	LOCALITE	(CAT)	COMPTE
B062	GOFFIN	72, r. de la Gare	Namur	B2	-3200
B112	HANSENNE	23, r. Dumont	Poitiers	C1	1250
B332	MONTI	112, r. Neuve	Genève	B2	0
B512	GILLET	14, r. de l'Eté	Toulouse	B1	-8700
C003	AVRON	8, r. de la Cure	Toulouse	B1	-1700
C123	MERCIER	25, r. Lemaître	Namur	C1	-2300
C400	FERARD	65, r. du Tertre	Poitiers	B2	350
D063	MERCIER	201, bvd du Nord	Toulouse		-2250
F010	TOUSSAINT	5, r. Godefroid	Poitiers	C1	0
F011	PONCELET	17, Clos des Erables	Toulouse	B2	0
F400	JACOB	78, ch. du Moulin	Bruxelles	C2	0
K111	VANBIST	180, r. Florimont	Lille	B1	720
K729	NEUMAN	40, r. Bransart	Toulouse		0
L422	FRANCK	60, r. de Wépion	Namur	C1	0
S127	VANDERKA	3, av. des Roses	Namur	C1	-4580
S712	GUILLAUME	14a, ch. des Roses	Paris	B1	0

PRODUIT			
NPRO	LIBELLE	RIX	QSTOCK
CS262	CHEV. SAPIN 200x6x2	75	45
CS264	CHEV. SAPIN 200x6x4	120	2690
CS464	CHEV. SAPIN 400x6x4	220	450
PA45	POINTE ACIER 45 (20K)	105	580
PA60	POINTE ACIER 60 (10K)	95	134
PH222	PL. HETRE 200x20x2	230	782
PS222	PL. SAPIN 200x20x2	185	1220

## 1.2 Utiliser une base de données

## 1.2 Utiliser une base de données

### Que peut-on faire de ces données ?

Avant tout, les conserver aussi longtemps que nécessaire !

Les interroger : *quel est le numéro, le nom et l'adresse des clients de Toulouse ?*

```
select NCLI, NOM, ADRESSE
from CLIENT
where LOCALITE = 'Toulouse';
```

facile

ou encore : *quelles sont les commandes des clients de Toulouse ?*

```
select NCOM
from COMMANDE
where NCLI in (select NCLI
               from CLIENT where LOCALITE = 'Toulouse');
```

un peu plus difficile

requêtes rédigées dans le langage SQL



## 1.2 Utiliser une base de données

### Que peut-on faire de ces données ?

**Vérifier une commande lors de son enregistrement : le client est-il connu ? son adresse a-t-elle changé ? les produits commandés sont-ils répertoriés ?**

**Produire les factures**

**Préparer le réapprovisionnement des produits en rupture de stock**

**Calculer le chiffre d'affaire mensuel**

**Etudier la répartition géographique des ventes**

**Et mille autres applications ...**

## 1.2 Utiliser une base de données

### Que peut-on faire de ces données ?

Un dernier exemple :

*calculer la répartition du chiffre d'affaire par localité et par produit*

```
select C.LOCALITE, P.NPRO, sum(D.QCOM*P.PRIX)
from   CLIENT C, COMMANDE M, DETAIL D, PRODUIT P
where  C.NCLI = M.NCLI and M.NCOM = D.NCOM
and    D.NPRO = P.NPRO
group by C.LOCALITE, P.NPRO;
```

**inutile d'essayer  
de comprendre  
pour l'instant !**

**Cette question complexe est résolue en une seule instruction SQL de 5 lignes !**

## 1.3 PREMIERES CONCLUSIONS

- Une **base de données** est constituée d'un ensemble de **tables**.
- Chaque **table** contient les données relatives à une **collection d'entités** de même nature.
- Chaque **ligne** d'une table reprend les données relatives à **une entité**.
- Chaque **colonne** d'une table décrit une **propriété commune** des entités.
- Les lignes d'une table sont **distinctes**. C'est sur une (voire plusieurs) colonne(s) que se joue cette unicité. Le jeu de colonnes dont les valeurs sont uniques constitue un **identifiant** de la table.
- Les lignes d'une table peuvent faire **référence** chacune à une ligne d'une autre table, où des informations sur une entité associée peuvent être obtenues. Les colonnes de référence s'appellent **clé étrangère** vers cette autre table.

## 1.3 Premières conclusions

- On évite d'enregistrer des données qu'il est possible de **calculer** à partir d'autres données enregistrées.
- On ne conserve pas dans une même table des données relatives à des entités de nature différente. On provoquerait en effet des **redondances** nuisibles.
- Le **langage SQL** permet de rédiger de manière simple et concise des *requêtes* d'extraction et de calcul de données de complexité quelconque.
- On ne construit pas une base de données pour satisfaire les besoins immédiats d'une application particulière. Une base de données est conçue pour **représenter le plus fidèlement possible** les informations relatives à une activité humaine, socio-économique, culturelle ou technique. Elle est **indépendante de ce qu'on fera de ces données**.
- La construction d'une base de données exige une **analyse rigoureuse** et attentive. Il est cependant possible à tout moment d'ajouter de nouvelles tables et de nouvelles colonnes afin de suivre l'évolution des besoins des utilisateurs.

# 1.4 LES SYSTEMES DE GESTION DE BASES DE DONNEES

**La gestion d'une base de données pose des problèmes complexes. Cette gestion est assurée par des logiciels spécialisés : les *systemes de gestion de bases de données* ou *SGBD*.**

**Organisation des données** : le SGBD organise les données en tables permanentes stockées sur disque; il crée les mécanismes garantissant un accès rapide aux données; il informe les utilisateurs sur ces structures.

**Gestion des données** : le SGBD garantit l'évolution cohérente des données; il vérifie que les contraintes (unicité, référence entre tables, etc.) sont respectées.

**Accès aux données** : le SGBD permet l'accès aux données à la fois par l'utilisateur occasionnel et par les programmes de traitement de données.

## 1.4 Les SGBD

**Protection contre les accidents** : le SGBD garantit l'intégrité et l'accessibilité des données en cas d'incident ou d'attaque.

**Gestion des accès concurrents** : le SGBD permet l'accès simultané aux données par des centaines voire des milliers d'utilisateurs. Il contrôle rigoureusement les opérations simultanées sur les mêmes données.

**Contrôle des accès** : le SGBD garantit que seuls les utilisateurs autorisés peuvent accéder aux données et les modifier.

## 1.5 LES DEFIS DES BD D'AUJOURD'HUI

**Multiplicité des types de données.** Une base de données moderne peut contenir des données multimédias, des données textuelles, des données spatiales (données GPS par exemple), des données historiques (plusieurs lignes par entité), des données semi-structurées.

**L'information incomplète.** Une base de données doit admettre qu'*on ne connaît pas tout sur tout* et que nos connaissances peuvent être erronées. Comment assurer au maximum la cohérence des données dans ce contexte ?

**Volumes et performances.** Une base de données peut contenir des dizaines de milliers de tables, des milliards de lignes. Comment garantir l'accessibilité de ces données, leur protection contre les incidents, des temps d'accès satisfaisants ?

## 1.5 Les défis des BD d'aujourd'hui

**Accès par des non informaticiens.** Malgré sa facilité d'utilisation (pour des requêtes simples), SQL est hors de portée de la plupart des utilisateurs non spécialistes. Des modes d'accès graphiques, en langage naturel ou par mode vocal, sont nécessaires pour en accroître l'usage.

**Maintenance et évolution.** La structure d'une base de données peut évoluer : ajouter ou supprimer une table, une colonne, une contrainte. Comment préserver les données et les programmes utilisateurs lors de cette évolution ?

**Les données distribuées et nomades.** Une base de données peut être répartie et/ou dupliquée sur plusieurs ordinateurs distribués géographiquement. Certains de ceux-ci peuvent être des appareils mobiles (embarqués, portables, *smart phones*). Comment garantir la cohérence, la protection et l'accessibilité des données ?



## 1.5 Les défis des BD d'aujourd'hui

**Les BD et le Web.** De nombreuses bases de données sont intimement liées au Web. En outre, le Web peut être vu comme une gigantesque base de données (largement incohérente et redondante !) Qu'en est-il des principes rigoureux des bases de données dans ce contexte ?

**Les données décisionnelles.** Les données ne servent pas seulement à contrôler la gestion et le fonctionnement d'une entreprise au jour le jour. Elles sont aussi souvent utilisées pour soutenir des décisions tactiques et stratégiques.

*Caractéristiques* : traitement de très gros volumes de données complexes pour produire une réponse courte (une heure de traitement de plusieurs téraoctets pour en extraire un seul nombre !). Comment structurer une base de données dans ce sens (= *entrepôts de données*) ?

## 1.6 UN PEU D'HISTOIRE

### Contenu

**1.1 Pré-condition : le disque magnétique**

**1.2 Les années 1960**

**1.3 Les années 1970**

**1.4 Les années 1980**

**1.5 Les années 1990**

**1.6 Les années 2000**

## 1.6 Un peu d'histoire - Pré-condition : le disque magnétique

Le **concept de base de données** est né dans les années 60, avec la généralisation des **disques magnétiques** permettant l'accès direct à une donnée, contrairement aux bandes magnétiques, limitées à l'accès séquentiel.



### Premier disque magnétique commercial : l'IBM 350, équipant le Ramac 305

- **Date** : de 1956 à 1969
- **Capacité** : 5 millions de mots de 7 bits (= 4,4 Mo)
- **Vitesse** : 1.200 rpm. Déplacement moyen : 0,6 sec (temps de commutation inconnu mais sans doute très important).
- **Géométrie** : 50 plateaux de 60 cm mais seulement 2 têtes. 100 cylindres, 100 pistes par cylindre. 5 secteurs par piste. 100 mots par secteur. 7 bits par mot.
- **Volume** : 2 m<sup>3</sup>. Poids : 1.075 kg
- **Nombre d'exemplaires fabriqués** : 100
- **Coût de location** : 35.000 \$/an (en 1956)

## 1.6 Un peu d'histoire - Pré-condition : le disque magnétique



### Mini-disque portable : 2 To

- **Date** : 2018
- **Capacité** : 2000 milliards d'octets (= 2 To)
- **Vitesse** : 5.400 rpm. Déplacement moyen : 0,016 sec.
- **Géométrie** : 1 plateau de 6,40 cm, 2 têtes
- **Volume** : 11 x 8 x 1,2 cm<sup>3</sup>. Poids : 90 gr.
- **Nombre d'exemplaires fabriqués** : 10.000.000 (estimation)
- **Coût d'achat** : 79 \$ (en 2018)

## 1.6 Un peu d'histoire - Les années 60

### Années 1960

- 1963 : Integrated Data Store (IDS) chez Honeywell-Bull. Développé par Ch. Bachman. Ancêtre des architectures standard de SGBD. Modèle de données dit **network**.
- 1965 : Information Management System (IMS) chez IBM. Au départ conçu pour North American Rockwell pour supporter le projet du 1er vol humain vers la lune. Initialement sur bande magnétique. Sera commercialisé en 1969 sous le nom IMS/360. Modèle de données dit **hierarchical**.
- 1968 : premier projet de recommandation CODASYL, inspirée d'IDS. Modèle de données dit **network**.

## 1.6 Un peu d'histoire - Les années 70

### Années 1970

- 1970 : article de E. F. Codd d'IBM définissant les principes des bases de données relationnelles (Codd, E., F., A Relational Model of Data for Large Shared Data Banks, in *Comm. ACM*, Vol. 13, No 6, June 1970).
- 1971 : première recommandation officielle CODASYL. Les rapports de 1973 et 1978 affineront et clarifieront les propositions. Origine de nombreux SGBD encore en activité aujourd'hui.
- 1973 : Première définition du langage SEQUEL, qui sera renommé SQL.
- 1974 : premier SGBD relationnel expérimental, le System/R d'IBM. Donnera naissance aux SGBD SQL/DS et DB2 dans les années 1980.
- 1975 : première version opérationnelle de INGRES, SGBD relationnel expérimental de l'université Berkeley. Langage QUEL concurrent de SQL. Donnera naissance à PostgreSQL. Des transfuges de l'équipe créeront Sybase.
- 1979 : première version commerciale d'un SGBD relationnel proposant SQL : Oracle.

## 1.6 Un peu d'histoire - Les années 80

### Années 1980

- 1982 : commercialisation de SQL/DS, SGBD relationnel d'IBM (machines MV).
- 1983 : commercialisation de DB2, SGBD relationnel d'IBM (machines MVS).
- 1987 : après 10 ans de travaux, premier standard effectif destiné aux SGBD relationnels.
- Les SGBD relationnels remplacent progressivement les SGBD traditionnels (CODASYL, IMS).
- Développement de SGBD expérimentaux admettant des valeurs structurées (tables *non plates*). Ces principes seront repris par les SGBD objet et relationnels objet.
- Développement de SGBD déductifs expérimentaux, basés sur les principes de déduction de la logique. Pas de succès commercial.
- Expérimentation sur les bases de données distribuées.

## 1.6 Un peu d'histoire - Les années 90

### Années 1990

- 1992 : Microsoft commercialise Sybase sous le nom **SQL Server**.
- Standard SQL2 ou SQL 1992. Introduit les notions de *primary key* et *foreign key*.
- Développement de SGBD expérimentaux **orientés objet**. Faible succès commercial mais certains de leurs principes seront repris par les SGBD relationnels objet.
- Apparition des SGBD relationnels objet.
- Microsoft envisage d'intégrer SQL Server comme composant système de Windows. Ce projet ne sera réalisé qu'avec Vista.
- Des versions légères sont proposées pour les systèmes mobiles et embarqués : cartes à puce, PDA, systèmes de capteurs, etc.



## 1.6 Un peu d'histoire - Les années 2000

### Années 2000

- 2000 : Apparition des SGBD gérant de manière native des documents **XML**. Faible succès commercial mais certains de leurs principes seront repris par les SGBD relationnels modernes.
- Standard SQL3 ou **SQL:1999**. Introduit notamment les requêtes récursives et les structures relationnelles objet.
- **SQL 2003**. Spécifie notamment les aspects multimédia, XML, modélisation spatiale et analyse de données (fouille de données, entrepôts de données).
- Les SGBD IMS et CODASYL issus des années 1960 sont toujours en activité !

## 1.7 ET ENSUITE ?

**Nous allons tenter de répondre à trois questions fondamentales :**

**Qu'est-ce qu'une base de données ?**

**Partie 1 - CONCEPTS DES BASES DE DONNEES**

**Comment utilise-t-on une base de données ?**

**Partie 2 - UTILISATION DES BASES DE DONNEES**

**Comment construit-on une base de données ?**

**Partie 3 - DEVELOPPEMENT DES BASES DE DONNEES**

**Fin du module 1**

**Module suivant :**  
***2. Concepts des bases de données***

