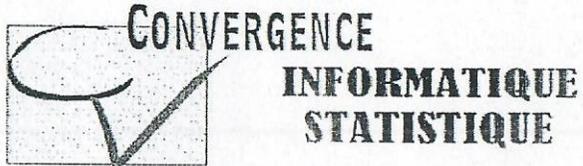


*Doublon*



- *Editeur de logiciels, Cartographie*
- *Maintenance, Réseaux, Formation*
- *Etudes statistiques, Enquêtes*
- *Fournitures de bureau, Consommables*
- *Matériel informatique, Accessoires*

**FORMATION A LA CONCEPTION  
ET A LA REALISATION D'UNE  
BASE DE DONNEES  
DECISIONNELLE**

-----  
**DATAWAREHOUSE - REPORTING -  
DATAMINING**

**Août 2009**

-----  
**CONVERGENCE Informatique Statistique SARI**  
Capital social 1 000 000 FCFA, RC SN-DKR-2008-B.1117, NINEA 28471792 .C2  
Zone B N° 01 / B App.06 – Tel/fax. (221) 33 824 20 37 – B.P. 14072 Dakar Peytavin – E mail : [convergencesa@orange.sn](mailto:convergencesa@orange.sn)  
RIB : Attijaribank K0052 01022 028451050182 / 24

# INFORMATIQUE DECISIONNELLE

## □ **Objectif global :**

Disposer d'un système d'information intégré permettant de réaliser des analyses sur un ensemble d'activités hétérogènes

## □ **Problématique :**

Devant la multiplication des sources de données à l'intérieur comme à l'extérieur de votre entreprise ou de votre organisation, il vous est de plus en plus difficile d'obtenir et de confronter les données qui vous intéressent et de les croiser dans un format compatible.

## □ **Solution :**

Une base de données fédératrice des informations à caractère décisionnelle au centre de notre système d'informations; qui doit nous permettre :

- d'homogénéiser l'ensemble des données en provenance des différentes sources,
- de centraliser ces données dans un endroit unique et accessible,
- de structurer ces données afin de les rendre utiles au pilotage de votre entreprise.

# PROBLEMATIQUE 1/3

## **1. Multiplicité des sources de données**

1. Jusqu'à une cinquantaine de sources différentes
2. Sources totalement autonomes
3. Aucune maîtrise du fonctionnement et de l'architecture de ces sources

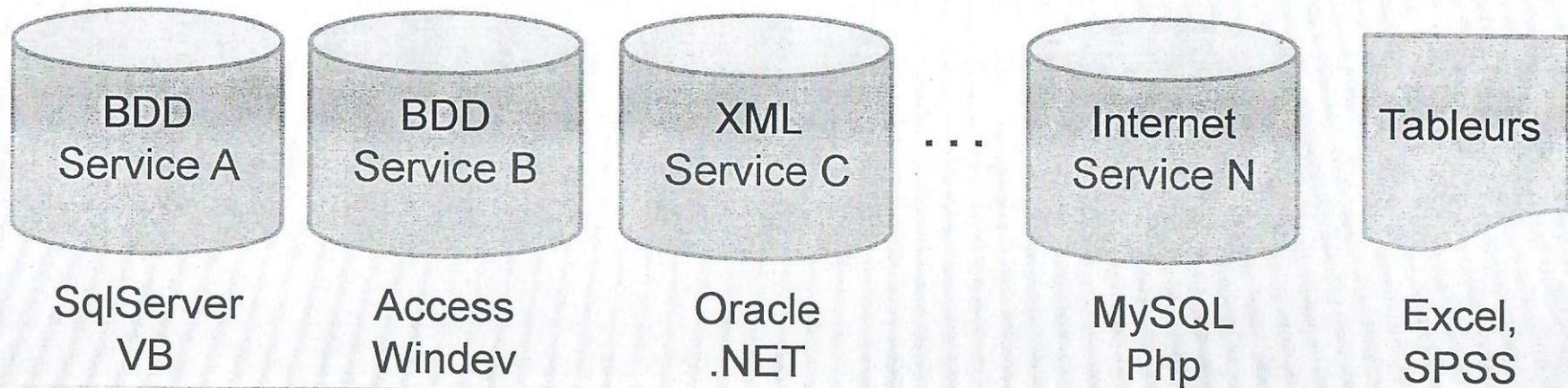
## **2. Diversité et Hétérogénéité des sources**

1. Bases de données – Fichiers tableurs plats – XML – Internet – ...
2. Access – Sqlserver – Oracle – MySql – ...
3. VB – Windev – .NET – Php – ...

## **3. Non harmonisation des sources**

1. Noms de tables synonymes différents
2. Codification de données différentes
3. Unités – Classification – Nomenclature différentes

# PROBLEMATIQUE 2/3



# PROBLEMATIQUE 3/3

REGION
01 - Dakar
02 - Diourbel
03 - Thiès
...

REGIONS
01 - Rg Dakar
03 - Rg Diourbel
01 - Rg Thiès
...

ZONE
DK - DAKAR
DL - DIOURBEL
TH - THIES
...

GEOGR
01 - Dakar
02 - Diourbel
03 - Thiès
...

COLONNE
Dakar
Diourbel
Thiès
...

UNITES
Km
Millions CFA
Tonnes
...

UNITES
m
CFA
Kg
...

UNITES
Kms
Milliers CFA
T
...

UNITES
Km, m
Millions CFA
Kg, Tonnes
...

UNITES
M
CFA
g, Kg, Tonnes
...



SqlServer  
VB

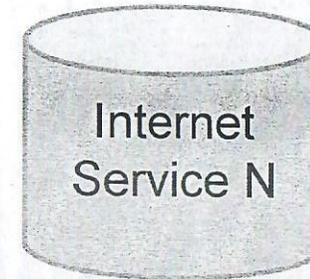


Access  
Windev

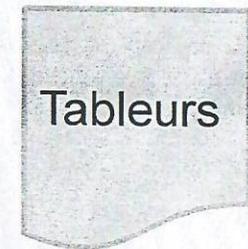


Oracle  
.NET

...



MySQL  
Php



Excel,  
SPSS

# ANALYSE DES SOURCES DE DONNEES 1/2

## 1. Interviews – analyse des rapports

1. Besoins des décideurs
2. Rapports d'activité et Etats financiers

## 2. Listing des informations demandées

1. Liste des informations quantitatives (sommables et non sommables)
2. Liste des informations qualitatives
3. Détermination des hiérarchies

Informations quantitatives INDICATEURS	Informations qualitatives AXES D'ANALYSE
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-

# ANALYSE DES SOURCES DE DONNEES 2/2

## 1. Matrice des sources de données

Pour chaque information demandée, établir :

1. Base de données, fichier tableur, ...
2. Table ou Feuille
3. Champ ou colonne
4. Méthode de calcul
5. Remarques

Information	Base de données, fichier tableur, ...	Table ou Feuille	Champ ou colonne	Méthode de calcul	Remarques
-					
-					
-					
-					
-					

# MODELISATION MULTIDIMENSIONNELLE 1/6

## □ Matrice de compatibilités

Pour chaque combinaison de (Indicateur X Axe d'analyse) établir si :

1. A été demandé
2. A un sens

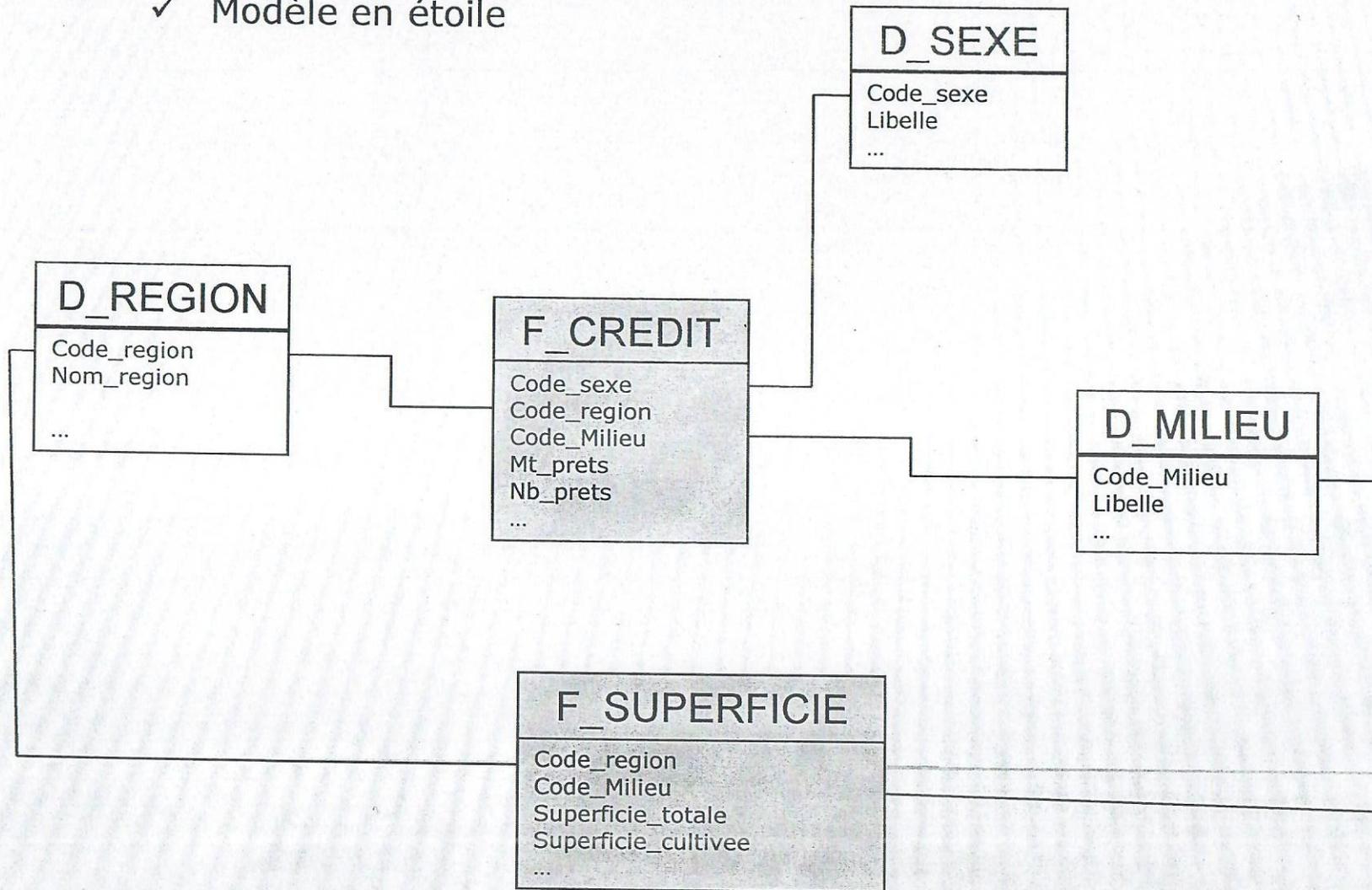
Informations qualitatives \ Informations quantitatives	Axe 01	Axe 02	Axe 03	Axe 04	Axe 05	Axe 06
Indicateur 01	X		X	X		
Indicateur 02		X	X			X
Indicateur 03		X				X
Indicateur 04						
Indicateur 05	X		X	X	X	
Indicateur 06	X	X	X		X	X
Indicateur 07	X				X	X
Indicateur 08		X				X
Indicateur 09						
Indicateur 10		X	X			X
Indicateur 11	X	X				X

**X** demandé  
**.** A un sens

# MODELISATION MULTIDIMENSIONNELLE 5/6

## □ Modèles multidimensionnels

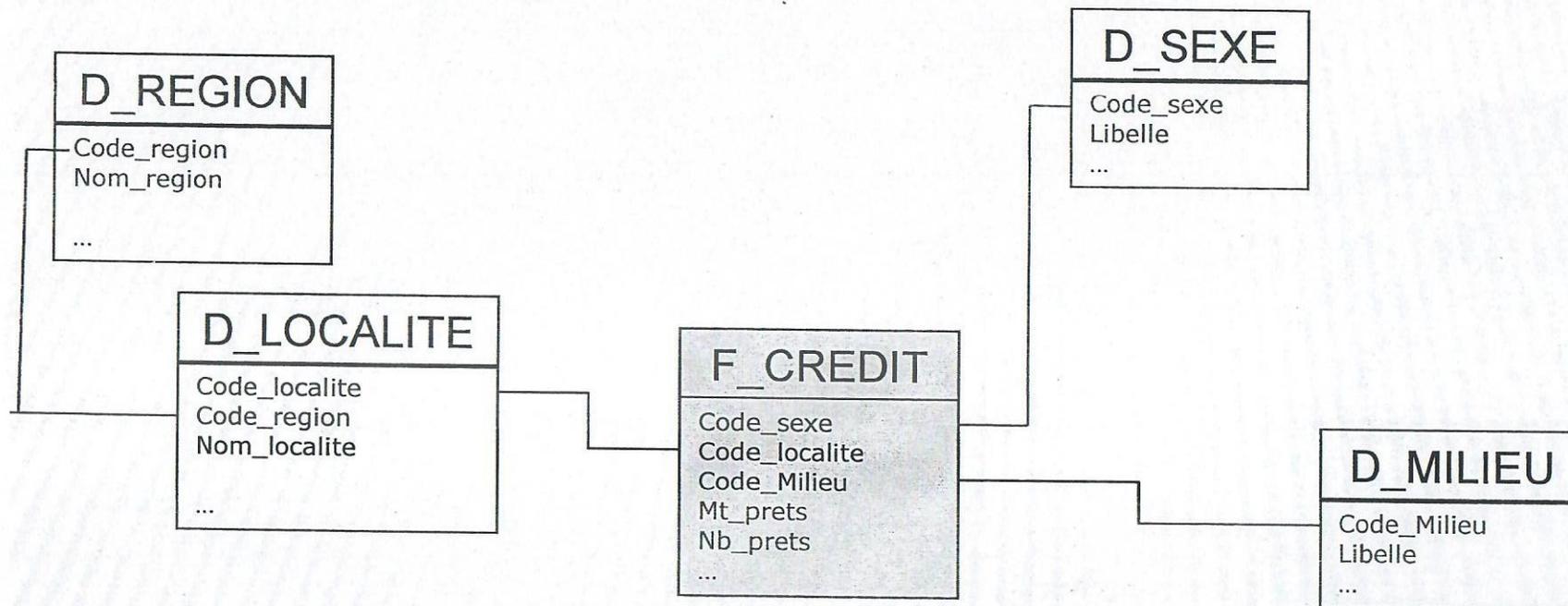
✓ Modèle en étoile



# MODELISATION MULTIDIMENSIONNELLE 6/6

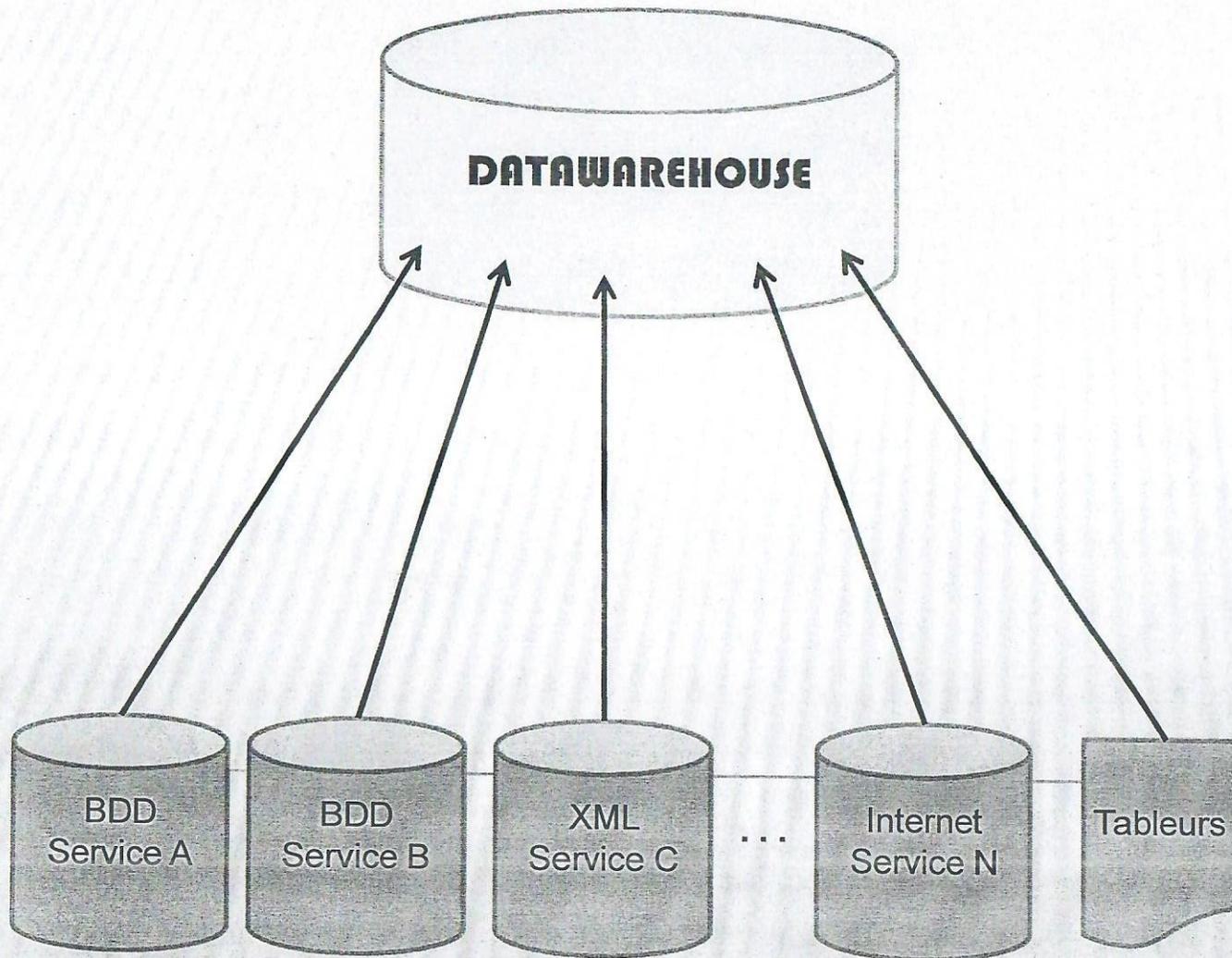
## ❑ Modèles multidimensionnels

- ✓ Modèle en flocon de neige



# ALIMENTATION DE L'ENTREPOT DE DONNEES 1/4

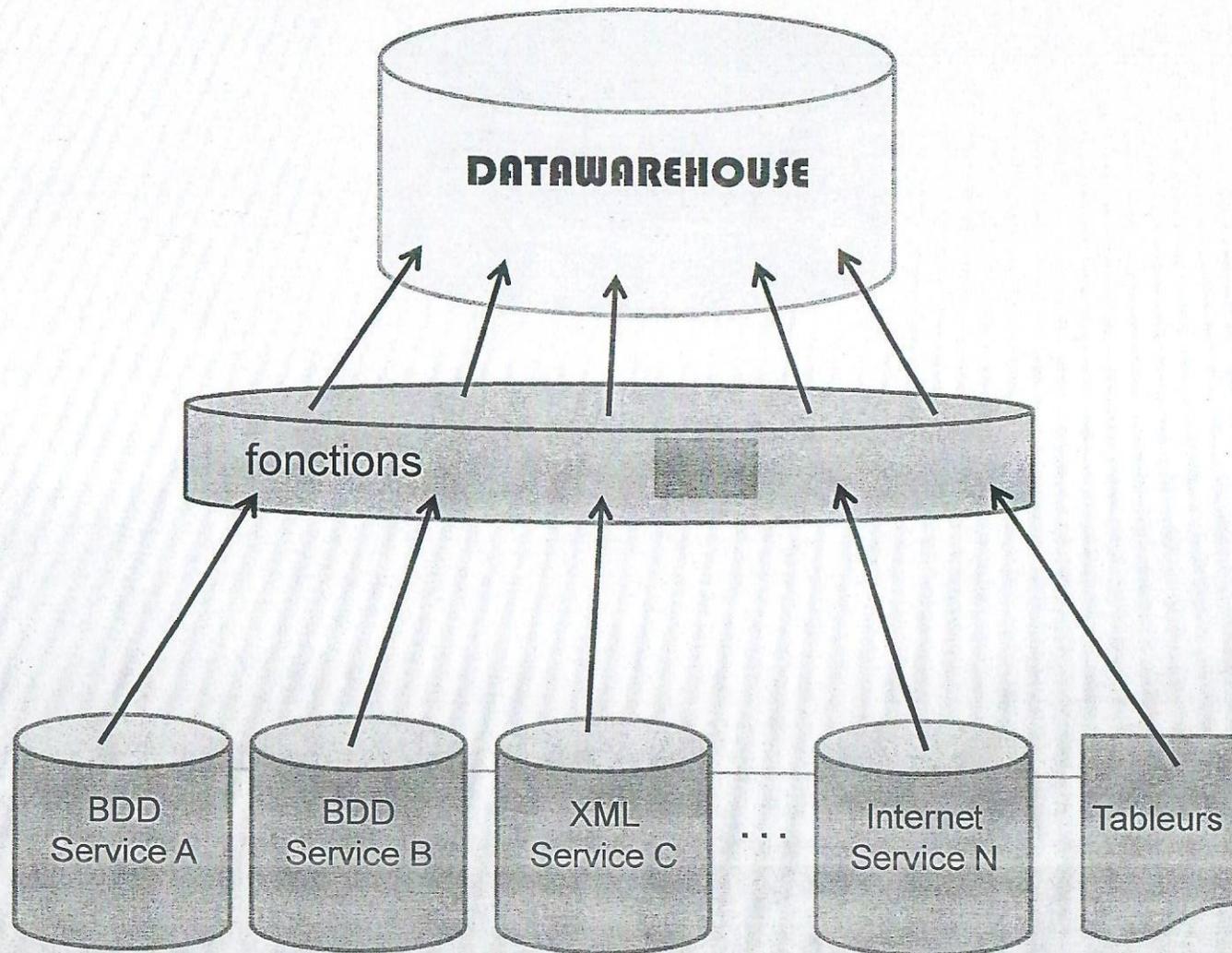
- Alimentation directe de l'entrepôt à partir des bases de données opérationnelles



*Les données directement chargées ne sont pas harmonisées donc impossibilité de réaliser des analyses croisées sur l'ensemble des données de l'entreprise*

# ALIMENTATION DE L'ENTREPOT DE DONNEES 2/4

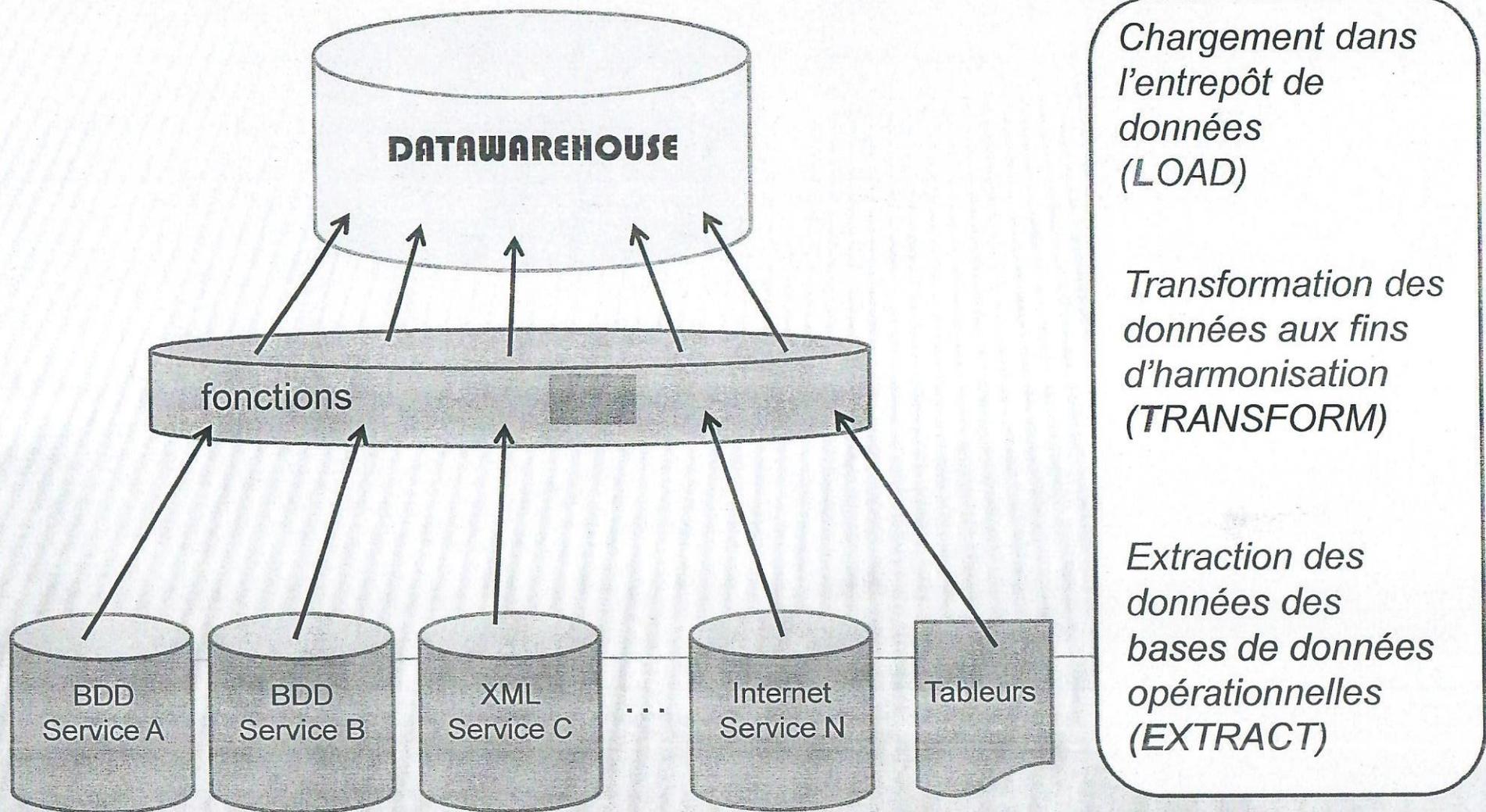
- ❑ Création d'une zone tampon d'harmonisation des données opérationnelles



*Les données sont extraites des bases opérationnelles ; provisoirement stockées dans la zone tampon avec des fonctions de transformation et des tables de correspondance*

# ALIMENTATION DE L'ENTREPOT DE DONNEES 3/4

## □ . E T L : Extraction – Transformation - Chargement



# ALIMENTATION DE L'ENTREPOT DE DONNEES 4/4

## □ Programmation de l'ETL

### ➤ Notion d'ETL

L'ETL est un logiciel à part entière chargé donc

- d'extraire les données à caractère décisionnel des bases opérationnelles
- de les harmoniser par des fonctions et procédures de transformation,
- de charger dans un entrepôt de données les données transformées

### ➤ Automatisation du chargement

L'ETL est caractérisé par

- Une périodicité ou une heure d'exécution
- La gestion des reprises en cas de panne ou d'erreur
- La non intervention humaine dans le processus d'ETL

# REPORTING DANS L'ENTREPOT DE DONNEES 2/4

## ❑ Interrogation multidimensionnelle de base

### ➤ Référentiel Indicateur / axe d'analyse

- Le système doit être flexible pour permettre de croiser n'importe quel indicateur avec les axes d'analyse associés sous la forme de tableau ou de tableau croisé
- Il doit être assez intelligent pour empêcher des requêtes insensées

### ➤ Filtres et tris

- Recherche multicritère : satisfaction de contrainte tant sur les indicateurs que sur les axes d'analyse sous forme de filtre
- Il permet de réaliser des tris tant sur les indicateurs que sur les axes d'analyse

### ➤ Classification

- Le système doit permettre de classer les résultats en groupes
- Il permet de regrouper les données suivant les axes d'analyse (ruptures) et d'effectuer des opérations d'agrégat (somme, moyenne, min, max, ...)

# REPORTING DANS L'ENTREPOT DE DONNEES 3/4

## □ **Graphiques – Scoring – Cartographie**

### ➤ Part de marché et tendances

- Le système doit permettre de réaliser des camemberts représentant la part de différentes modalités part rapport a un indicateur
- Le système doit permettre de réaliser des histogrammes représentant dans le temps l'évolution d'un indicateur

### ➤ Top N et Bottom N

- Le système permet de déterminer les N meilleurs ou les N plus faibles ou de transformer du quantitatif en qualitatif

### ➤ Systèmes d'information géographiques

- Le système doit permettre de représenter dans une carte la distribution d'un indicateur suivant plusieurs zones géographiques

# REPORTING DANS L'ENTREPOT DE DONNEES 4/4

## □ Data Mining

### ➤ Panier de la ménagère

- Permet de réaliser un profilage du comportement des individus sur la base de la répétition d'association

### ➤ Corrélations et recherche d'influence

- Permet de mettre en évidence la causalité entre des indicateurs et des axes d'analyse

### ➤ Détection de saisonnalité et techniques de recherche de fraude

- Permet de déceler dans un ensemble de données des répétitions à intervalle de temps réguliers ou au contraire des données exceptionnelles

### ➤ Apprentissage supervisé et non supervisé

- Permet de prévoir le comportement futur des individus sur la base du comportement d'une population d'individus semblables