

- 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 S.A.R.L.
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 Poyavin - E mail : convergenceca@orange.sn
RIB : Attijari bank 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

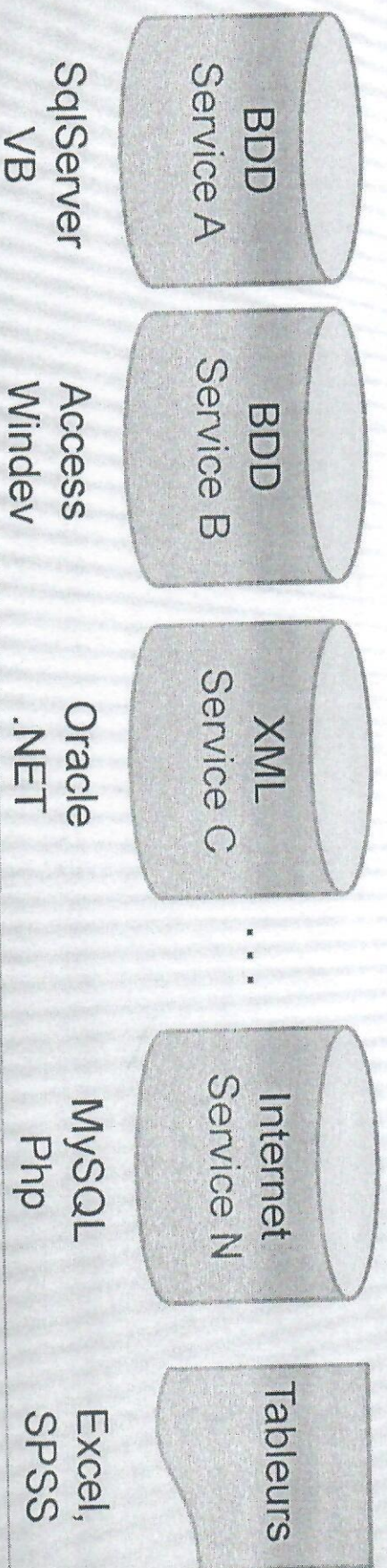
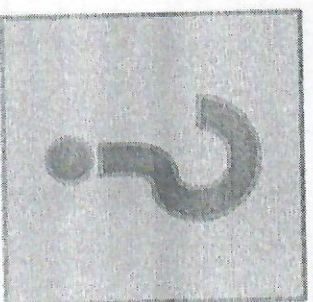
2. Diversité et Hétérogénéité des sources

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



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
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-

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	Axe 01	Axe 02	Axe 03	Axe 04	Axe 05	Axe 06
Informations quantitatives						
Indicateur 01	X		X	X	.	X
Indicateur 02		X	X			X
Indicateur 03	.	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 2/6

□ Tables de dimensions

1. Les axes d'analyse deviennent des tables de dimension
2. Ces dimensions ont un code et un libellé
3. Seuls les attributs totalement dépendants, (même dans le temps) y sont ajoutés

D_REGION
Code_region
Nom_region
...

D_SEXE
Code_sexe
Libelle
...

D_MILIEU
Code_Milieu
Libelle
...

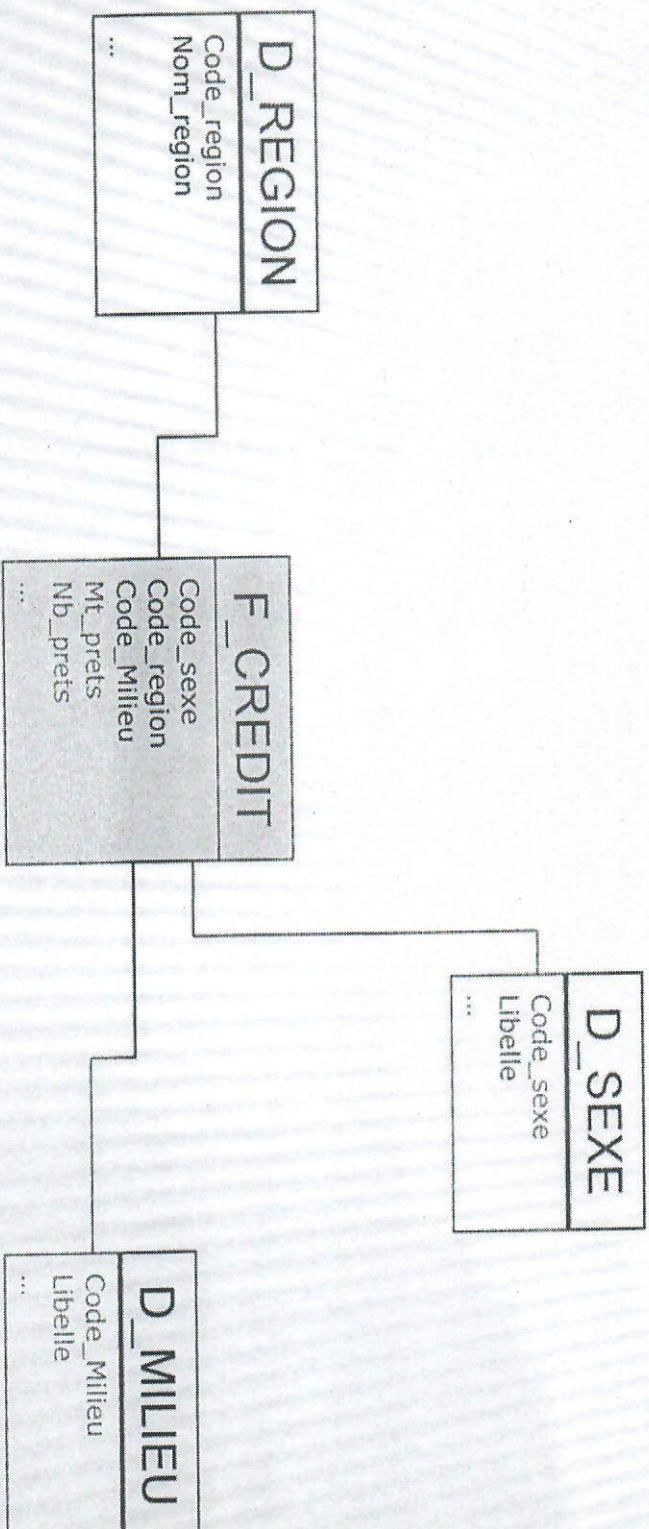
MODELISATION MULTIDIMENSIONNELLE 3/6

□ Regroupement d'indicateurs selon les axes d'analyse

1. Cas d'indicateurs isolés
 - ✓ Chaque indicateur est associé à une table contenant l'ensemble des dimensions qui lui sont associées
2. Optimisation des tables de données
 - ✓ Les faits qui partagent exactement les mêmes dimensions sont optimisées dans la même table
3. Tables de faits
 - ✓ On appelle table des faits les tables issues de cette optimisation

MODELISATION MULTIDIMENSIONNELLE 4/6

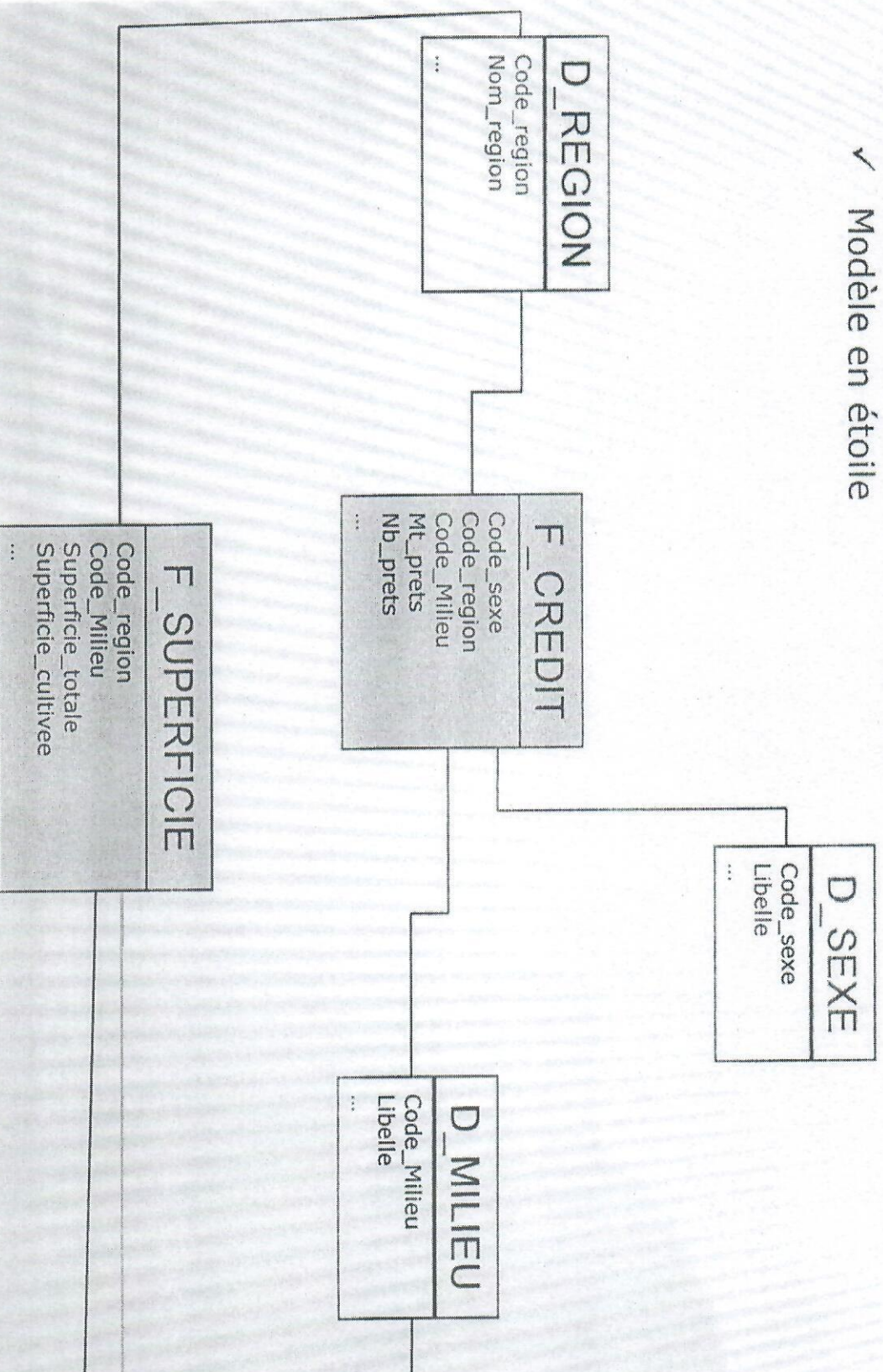
❑ Tables de faits



MODELISATION MULTIDIMENSIONNELLE 5/6

❑ Modèles multidimensionnels

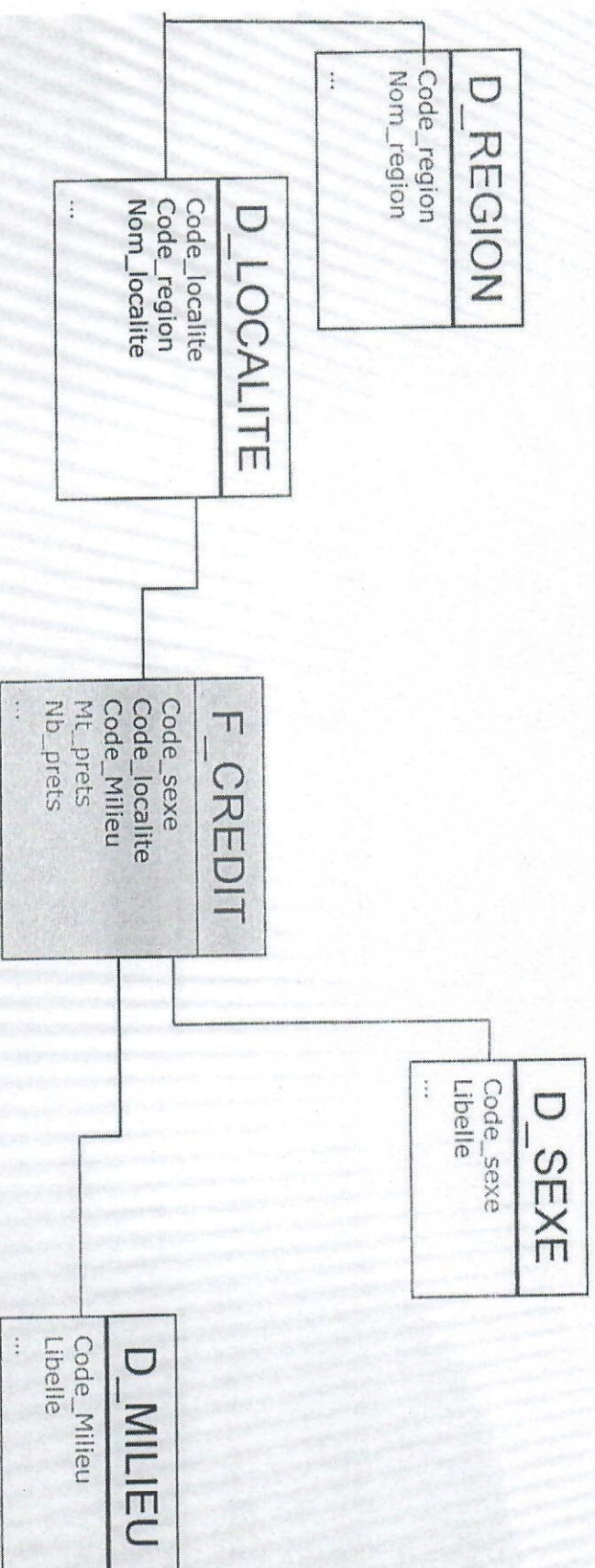
✓ Modèle en étoile



MODELISATION MULTIDIMENSIONNELLE 6/6

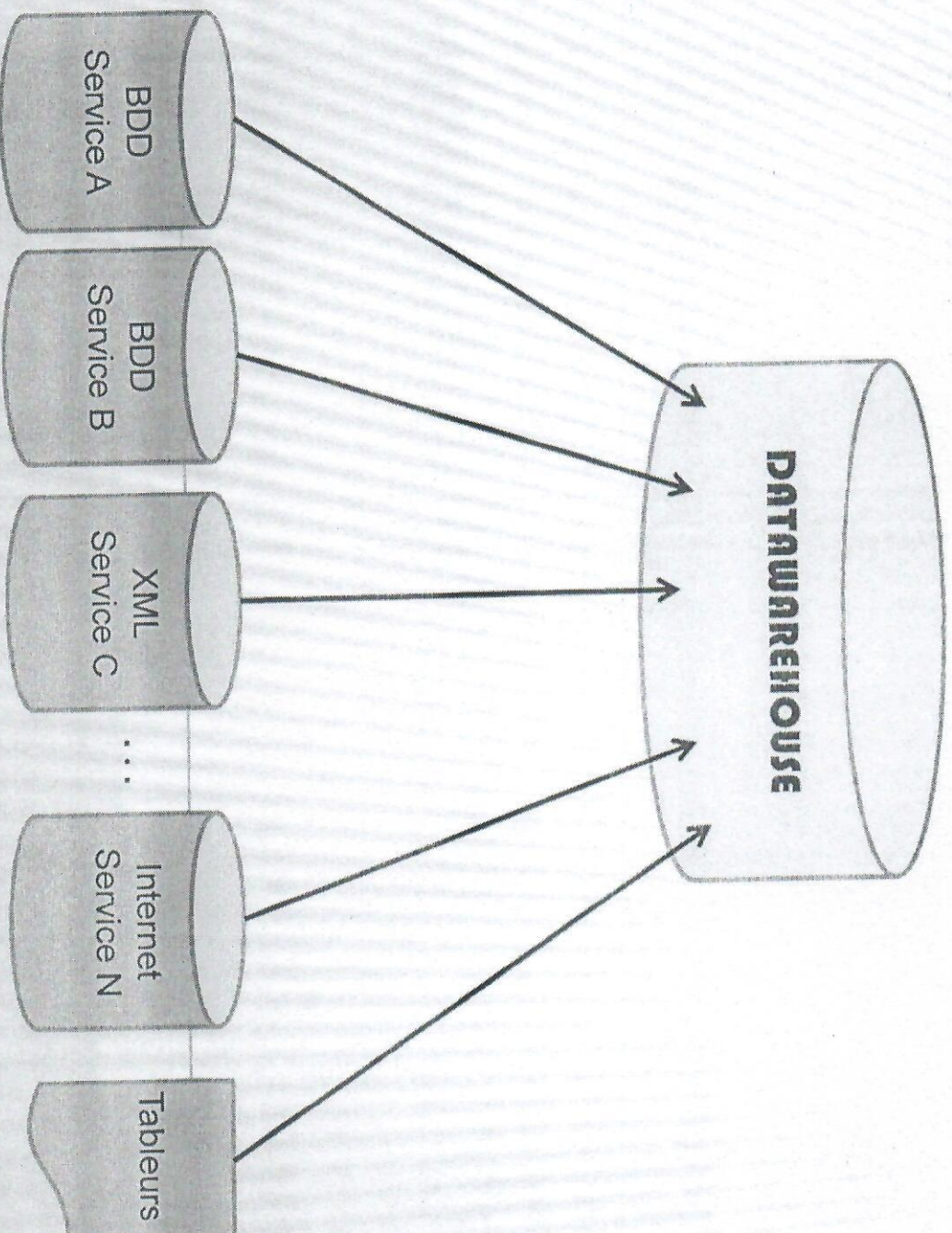
❑ Modèles multidimensionnels

✓ Modèle en flocon de neige



ALIMENTATION DE L'ENTREPOT DE DONNEES 1/4

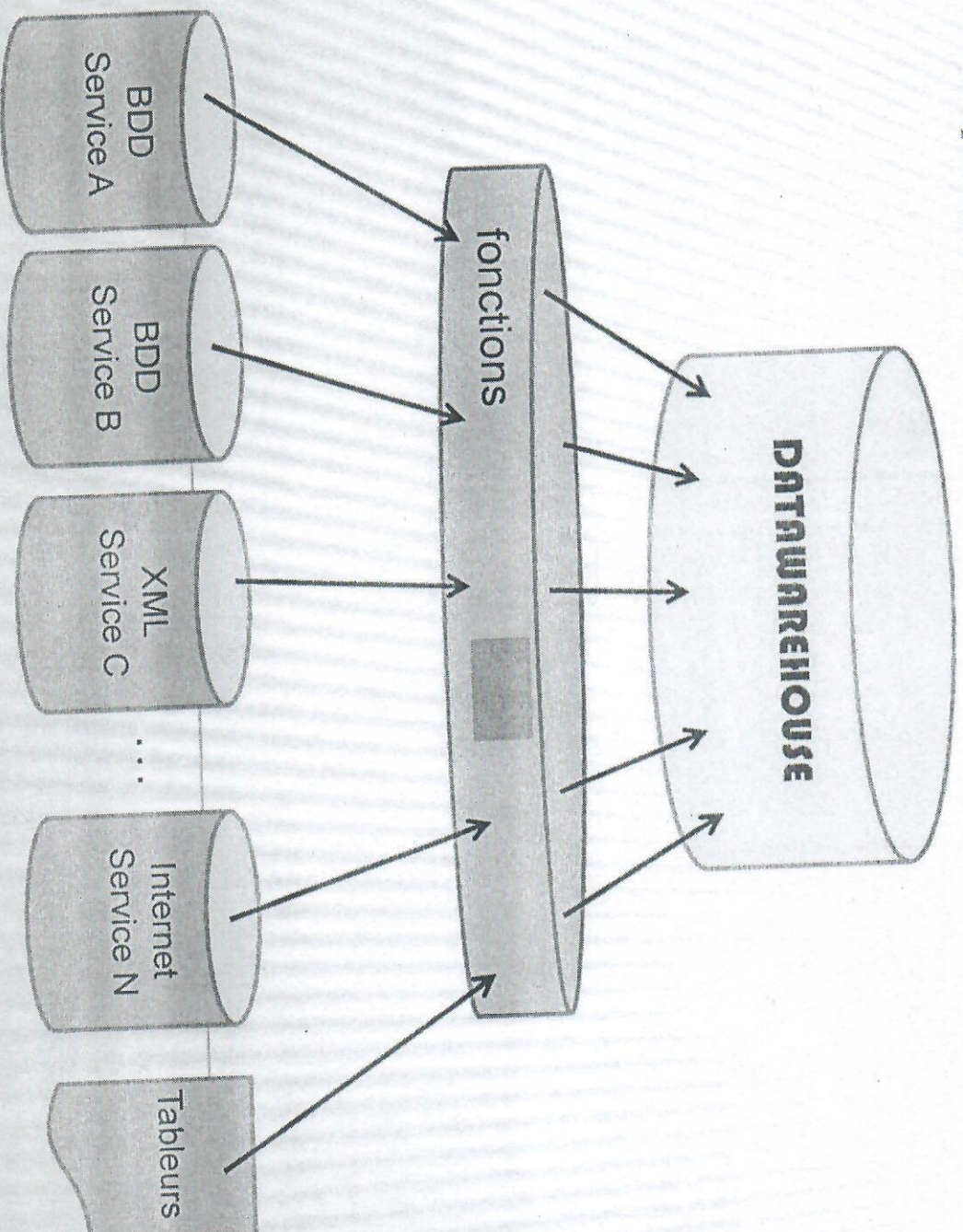
- Alimentation directe de l'entrepôt à partir des bases de données opérationnelles



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

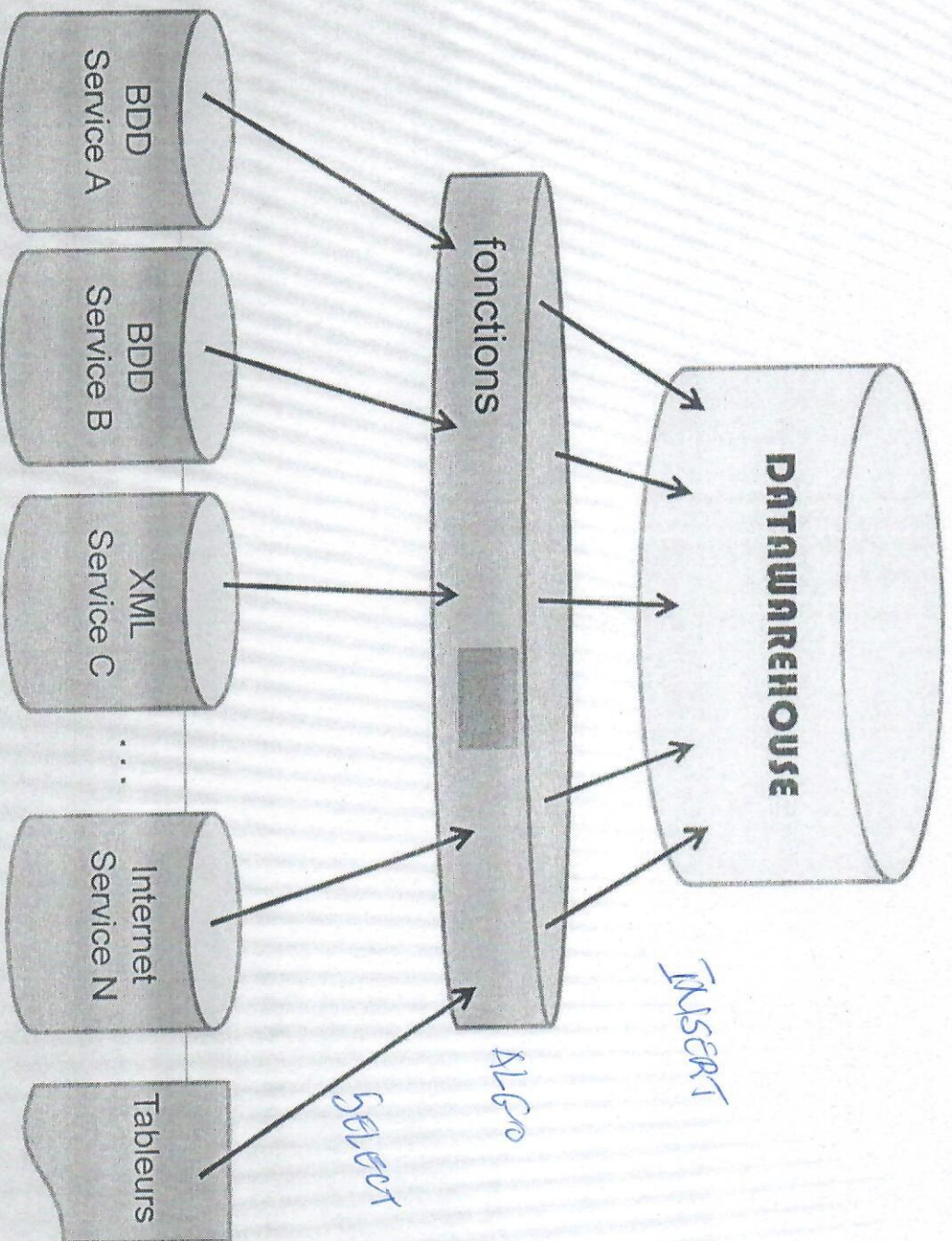
- Création d'une zone tampon d'harmonisation des données opérationnelles



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



Chargement dans l'entrepôt de données (LOAD)

Transformation des données aux fins d'harmonisation (TRANSFORM)

Extraction des données des bases de données opérationnelles (EXTRACT)

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 1/4

□ Vue d'ensemble d'un outil de reporting

- Le système doit être ergonomique et intuitif
- Aucune notion SQL ou autre langage informatique ne doit être requise pour exploiter la solution
- Une documentation et une aide contextuelle doivent être fournies et disponibles
- Un système de mouchard permet de retracer les requêtes des utilisateurs pour disposer du profil des besoins évolutifs des décideurs
- Un système de sécurité permet à l'administrateur du système (qui peut ne pas être un informaticien) de restreindre l'interrogation de certains indicateurs et de certains axes d'analyse à certains utilisateurs

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 par rapport à 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