

Bases de Données Hétérogènes

Didier DONSEZ

Université Joseph Fourier

IMA – IMAG/LSR/ADELE

`Didier.Donsez@imag.fr`

Plan

- Introduction
- Architecture de SGBD-D Hétérogènes
- Distribution - Hétérogénéité - Autonomie
- Problèmes à Résoudre

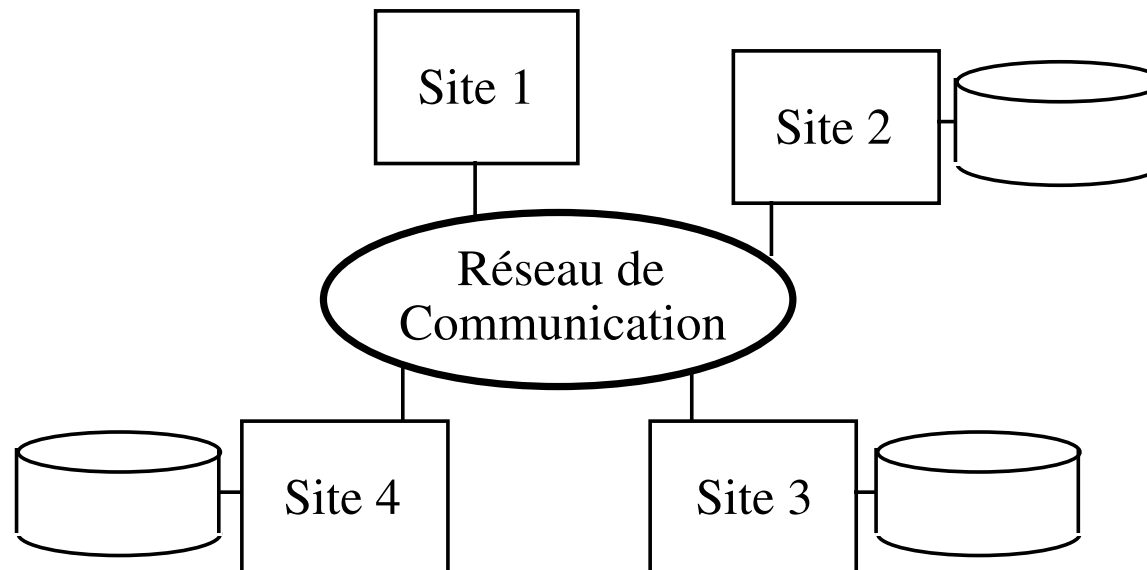
Rappel

■ BD Distribuée

- une collection de BDs logiquement reliées et distribuées entre plusieurs sites

■ SGBD Distribué

- gère une BD-D et fournit les mécanismes d'accès rendant la distribution transparente à l'utilisateur.



Les Besoins des Applications

■ Applications traditionnelles (gestion)

- types de données simples
- données structurées
- accès rapide par ensemble
- partage et sécurité
- ⇨ Bien supporté par les SGBDs relationnels

■ Nouvelles Applications (SID, CAO, IA, Médical, ...)

- données complexes
- manipulation complexe
- règles
- autre : schéma dynamique, multimédia, déclencheur, version
- ⇨ Mal supporté par le relationnel

Les Développements Technologiques

■ Amélioration des communications

- Réseaux Locaux plus rapide
FDDI, Fiber Channel, Ethernet 100
- Réseaux longues distances plus rapides, plus surs, plus Etendus
Internet, lignes T1, RNIS, ATM

■ Amélioration des stations de travail

- Meilleur prix/performance
- amélioration des possibilités - station multiprocesseur

Les Ilots d'Information de l'Entreprise

■ Développement des BDs indépendantes

- différentes machines : mainframes, minis, micros, stations de travail
- différents SGBDs : hiérarchique, relationnel, objet
- Besoins spécifiques des départements

■ Conséquences

- Redondance des données et inconsistance
- difficulté de localiser les données “utiles” et de les intégrer rapidement

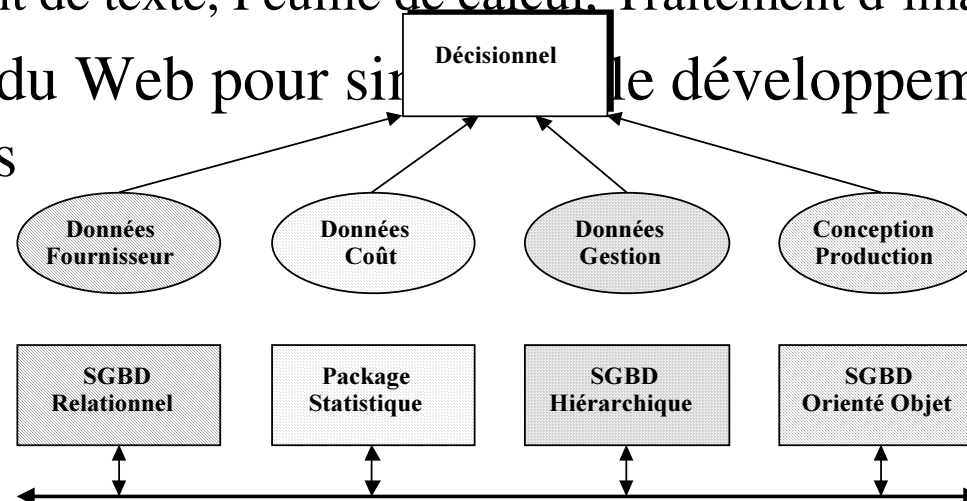
■ Données “Utiles”

= consistantes, à jour, et facilement accessibles

- ⇒ Intégration logique des données distribuées et hétérogènes de l'entreprise

L'Environnement Futur

- Relationnel \Rightarrow applications traditionnelles
- Modèles sophistiqués \Rightarrow applications avancées
- Héritage d'anciens systèmes (legacy systems)
 - 45% des SGBDs sur Mainframe sont IMS (hiérarchique)
 - De nombreuses applications utilisent encore des SGFs (VSAM)
- Intégration des systèmes "non-gestionnaires" de données
 - Traitement de texte, Feuille de calcul, Traitement d'images.
- Utilisation du Web pour le développement des applications



La Demande pour l'Interopérabilité entre SGBDs

■ Conséquence de la Distribution, des Réseaux et des SGBDs

- 3 développement de BDs indépendantes par différents groupes d'utilisateurs
- 3 survivance des BDs existantes
- 3 accès aux BDs par différents outils

■ Difficultés

- de connaître l'utilité et la localisation d'une donnée
- d'intégrer rapidement des données hétérogènes
- de combiner l'activité d'outils hétérogènes

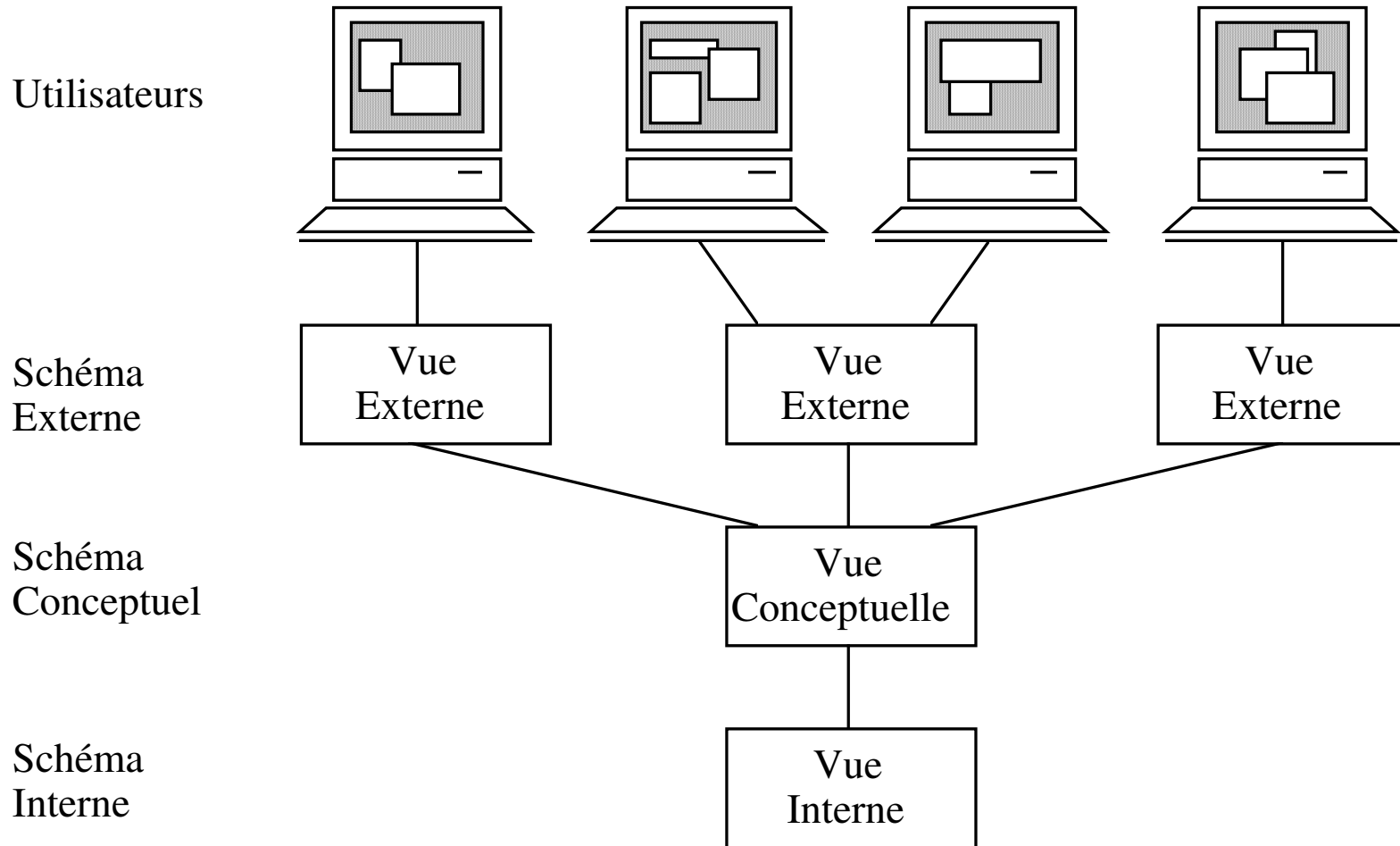
■ Interopérabilité

- ⇨ interaction intelligente entre des systèmes hétérogènes
 - ex : SGBD/feuille de calcul
- ⇨ combinaison des SGBDs, des systèmes répartis, des traitements de textes, Ö

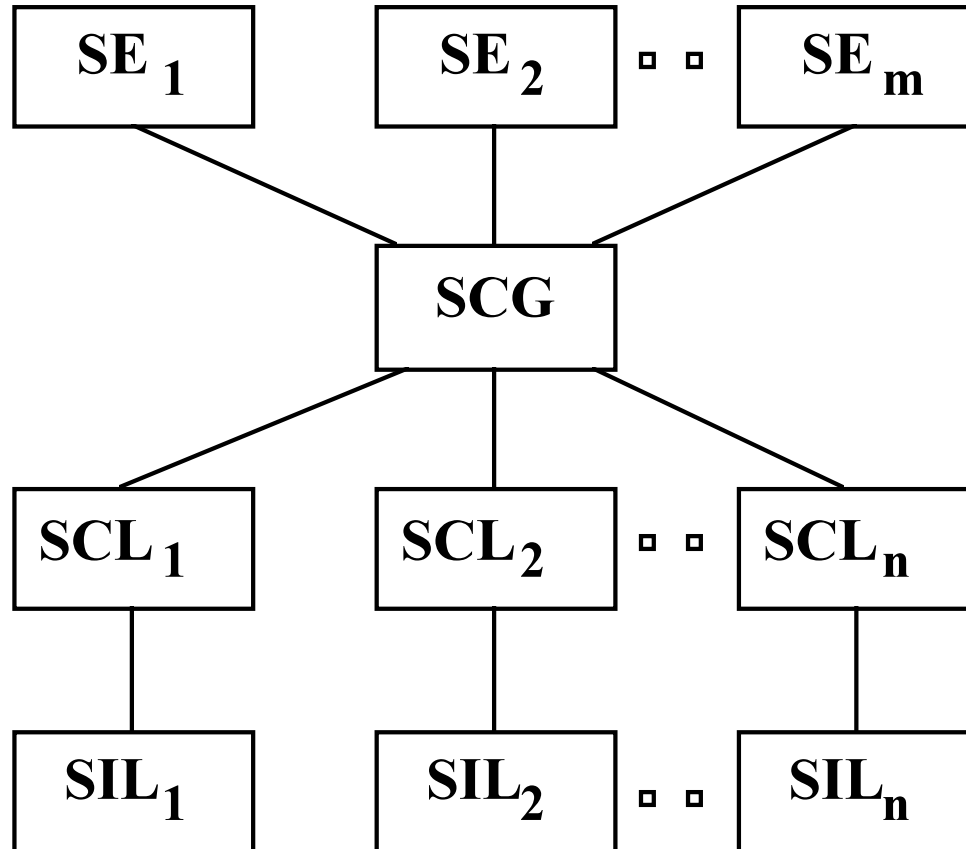
Les Objectifs des SGBDs Interopérables

- Interopérabilité des BDs existantes, distribuées et hétérogènes au sien d'un Multi-SGBD
- intégration de types de données conventionnels ou non-conventionnels comme dans un SGBD-OO
- Interopérabilité avec des applications existantes
 - Héritage de l'existant
- Utilisation des composants logiciels existants
 - extensibilité par la définition de méthodes d'objets
- Invoquer arbitrairement une application et créer des combinaisons d'invocation

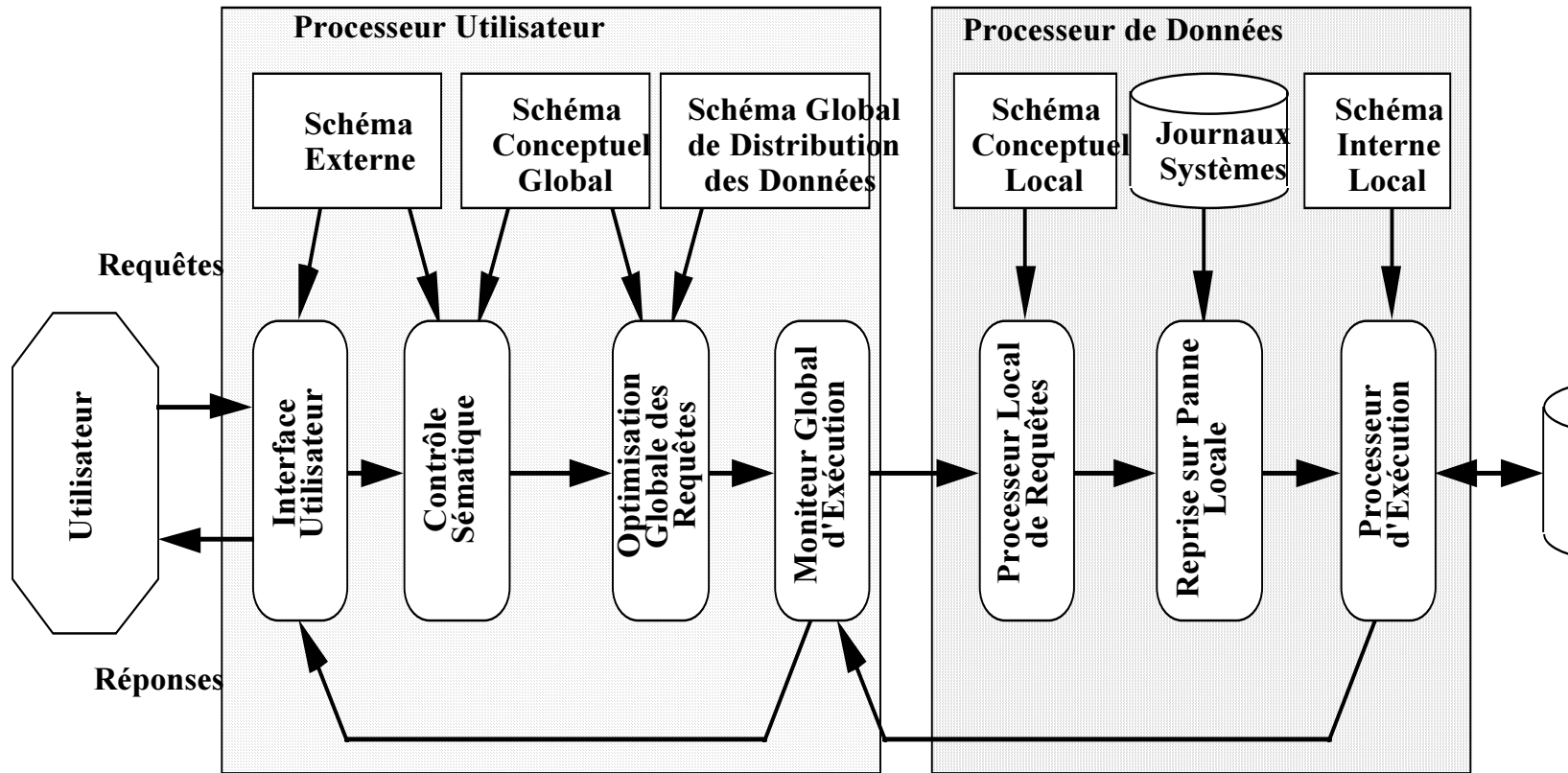
L'architecture ANSI/SPARC



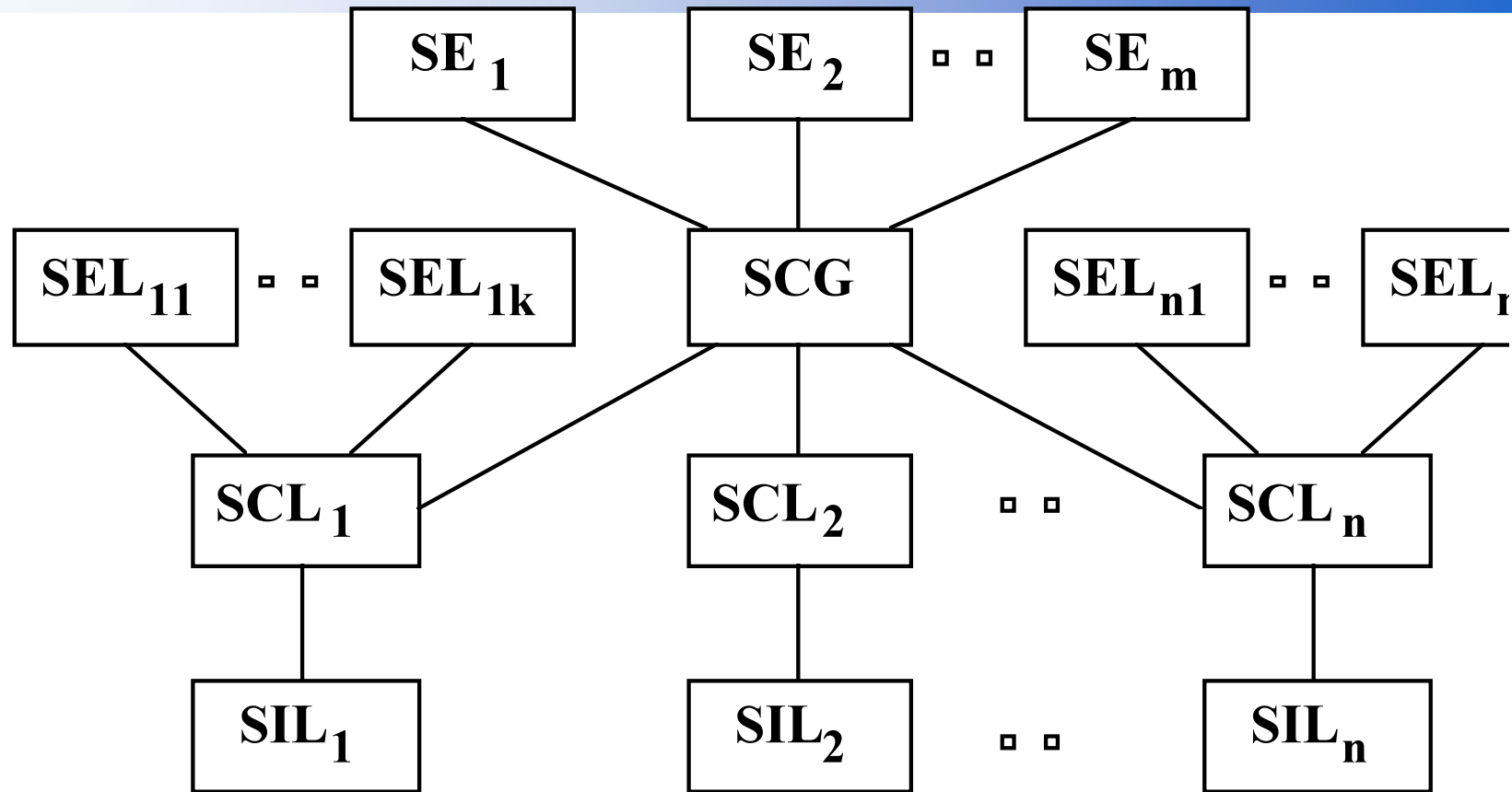
L'architecture logique d'un SGBD-D



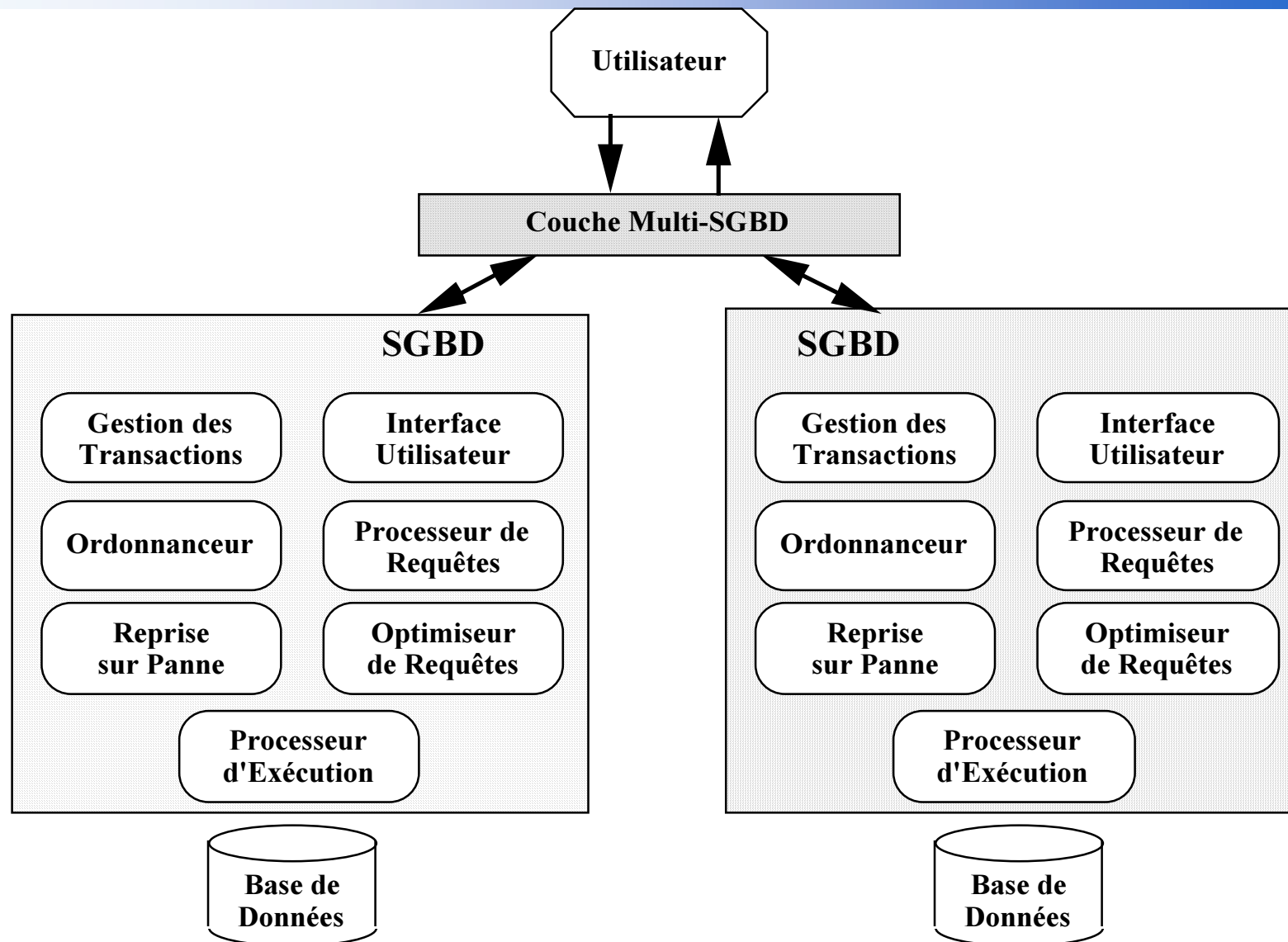
L'architecture physique d'un SGBD-D



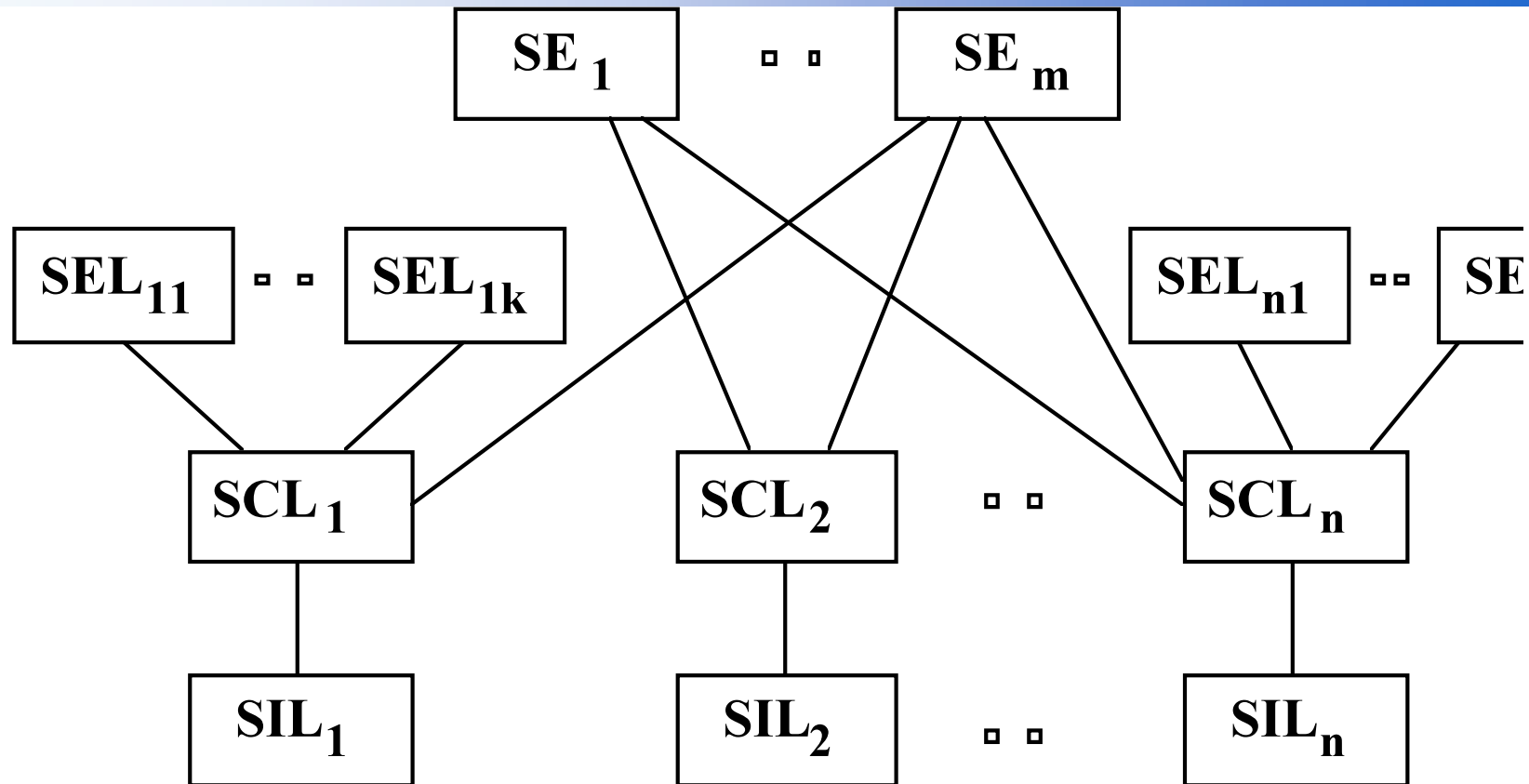
L'architecture logique de Multi-SGBD



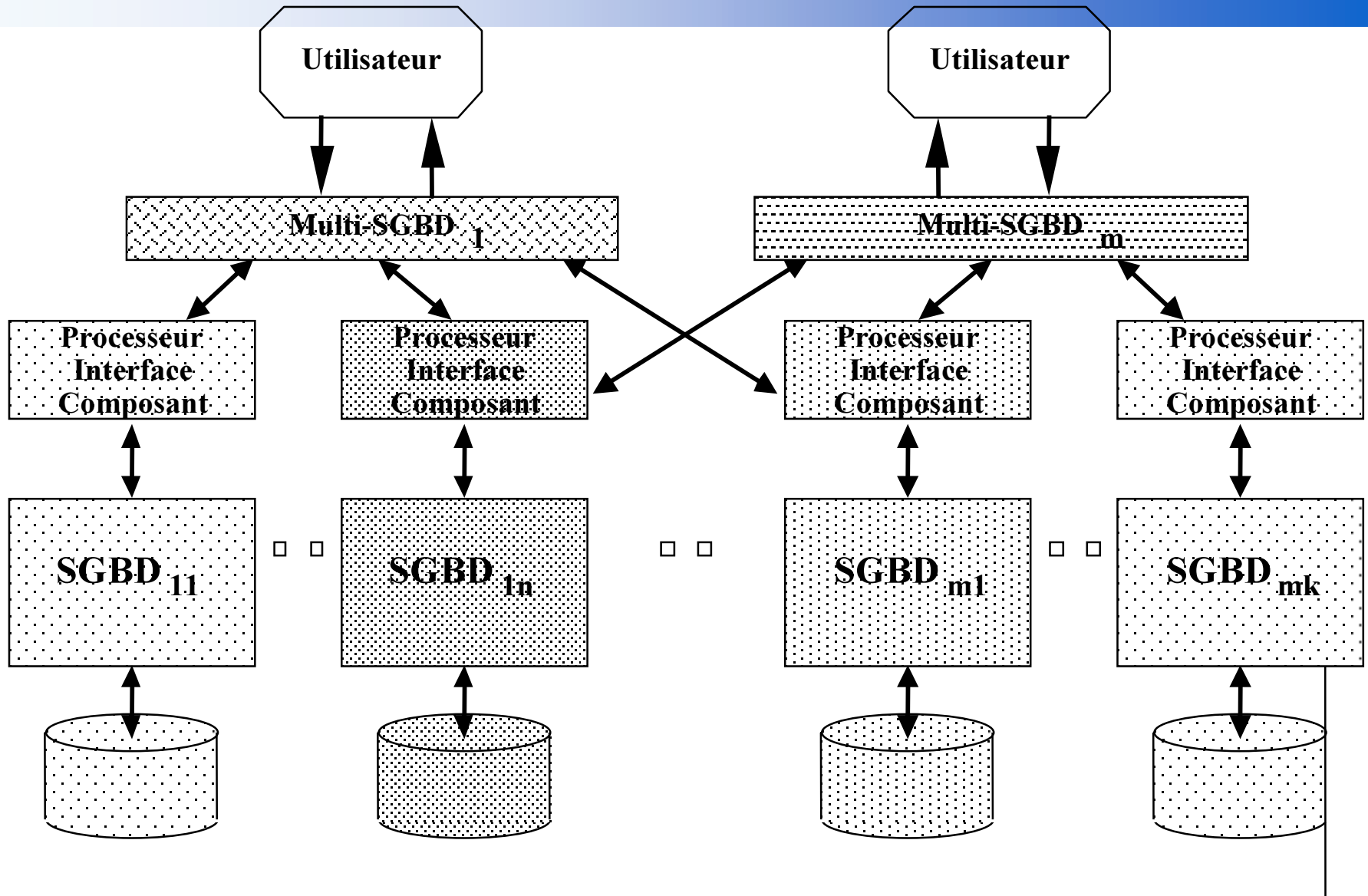
L'architecture physique d'un Multi-SGBD



L'architecture logique de Multi-SGBD - sans SCG



L'architecture physique d'un Multi-SGBD sans SCG



Les Dimensions du problème

■ Distribution

- bien maîtrisé

■ Hétérogénéité

- Plusieurs degrés

■ Autonomie

- Difficile

Les différents degrés d'Hétérogénéité

- chaque BD est gérée par le même SGBD sur des systèmes différents
 - ex: ORACLE sur UNIX, VMS, DOS

- chaque BD est gérée par un SGBD différent mais avec le même modèle
 - ex: ORACLE, DB2, INGRES

- chaque BD est gérée par un SGBD différent avec un modèle de données différent
 - ex: DB2 et IMS (problème pratique pour IBM)

- chaque BD est une source de données structurées ou non

Les différentes Formes d'Autonomie

■ Autonomie de Conception

- un SGBD décide d'une conception reliée à ses propres besoins
 - données, représentation, interprétation, implantation, Ö
- appelé Autonomie physique

■ Autonomie de Communication

- un SGBD décide de comment et avec quel autre SGBD il doit communiquer

■ Autonomie d'Exécution

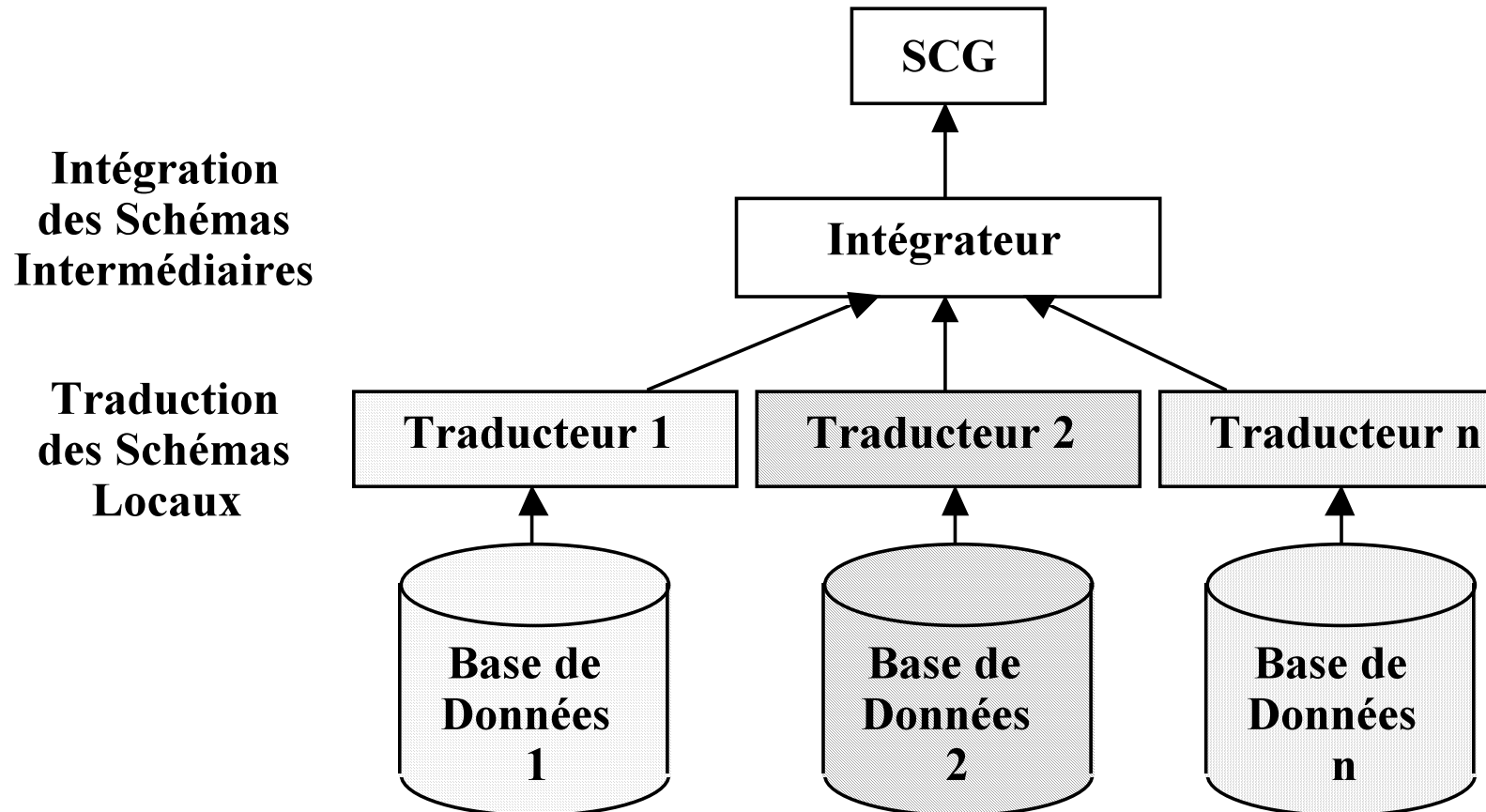
- un SGBD peut exécuter des opérations localement de la manière qu'il veut
- appelé Autonomie de Site

Les problèmes à Résoudre

- Intégration de Bases de Données
- Calcul des Requêtes
- Gestion des Transactions

intégration de Bases de données

- c'est le processus d'intégration conceptuelle des informations provenant des différentes bases composantes



intégration de schéma

■ Pré-intégration

- identification des éléments reliées
 - ex: domaines Equivalents
- spécification des règles de conversion
 - ex: 1 pouce = 2,54 cm
 - ex: salaire en \$ <-> DM <-> £ <-> FF

■ Comparaison

- conflits de nommage
 - synonymes, homonymes
- conflits structurels
 - types, clés, dépendances

■ Conformité

- résolution des conflits de noms et structurels

■ Fusion et Restructuration

- fusion des schémas intermédiaires
- fournir le “meilleur” schéma intègre

Calcul des requêtes

■ La complexité du calcul des Requêtes est élevée

- variations d'un SGBD \ddagger l'autre
 - pour les possibilités
 - pour les coûts
 - pour l'optimisation
- difficulté de déplacer les données entre SGBDs

Gestion de Transactions

- plusieurs transactions ‡ coordonner
 - Terminaison des transactions, contrôle de Concurrences

■ Moniteurs Transactionnels

- OSI/TP, X/Open, OMG, ...
- Tuxedo, ...