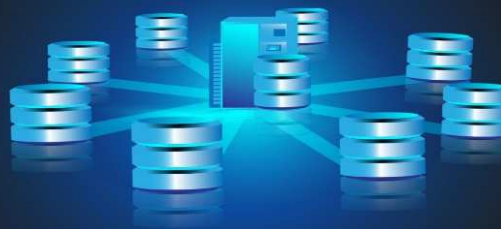


Data Warehouse

Modélisation Logique (OLAP)



Modélisation Logique – Données multidimensionnelles

- Données perçues à travers plusieurs dimensions
- Données qualifiées de **multidimensionnelles**, indépendamment de leur support
 - tables relationnelles
 - tableaux multidimensionnels

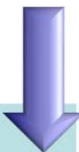


Produit	Region	Ventes
Clous	EST	50
Clous	OUEST	60
Clous	CENTRE	100
Vis	EST	40
Vis	OUEST	70
Vis	CENTRE	80
Boulons	EST	90
Boulons	OUEST	120
Boulons	CENTRE	140
Nettoyeurs	EST	20
Nettoyeurs	OUEST	10
Nettoyeurs	CENTRE	30



Modélisation Logique – Données multidimensionnelles

- Données perçues à travers plusieurs dimensions
- Données qualifiées de **multidimensionnelles**, indépendamment de leur support
 - tables relationnelles
 - tableaux multidimensionnels



	Est	Ouest	Centre
Clous	50	60	100
Vis	40	70	80
Boulons	90	120	140
Nettoyeurs	20	10	30

Produit	Region	Ventes
Clous	EST	50
Clous	OUEST	60
Clous	CENTRE	100
Vis	EST	40
Vis	OUEST	70
Vis	CENTRE	80
Boulons	EST	90
Boulons	OUEST	120
Boulons	CENTRE	140
Nettoyeurs	EST	20
Nettoyeurs	OUEST	10
Nettoyeurs	CENTRE	30



Modélisation Logique – Données multidimensionnelles

- Requêtes décisionnelles de type :
 - Quel est le total des ventes des « clous » dans la région Ouest ?
- Tables relationnelles : on peut traiter quelques centaines de tuples par seconde
- Tableau multidimensionnel : on peut rajouter en lignes et en colonnes plus de 10 000 valeurs par seconde

 Accélérer les temps de réponses



Préparer/pré-calculer des sous totaux

51



Produit	Region	Ventes
Clous	EST	50
Clous	OUEST	60
Clous	CENTRE	100
Clous	Total	210
Vis	EST	40
Vis	OUEST	70
Vis	CENTRE	80
Vis	Total	190
Boulons	EST	90
Boulons	OUEST	120
Boulons	CENTRE	140
Boulons	Total	350
Nettoyeurs	EST	20
Nettoyeurs	OUEST	10
Nettoyeurs	CENTRE	30
Nettoyeurs	Total	60
Total	EST	200
Total	OUEST	260
Total	CENTRE	350
Total	Total	810



- Calcul des totaux: 28 accès en lecture et 8 accès en écriture
- Un SGBDR lit 200 enreg/s et en écrit environ 20/s.



Produit	Region	Ventes
Clous	EST	50
Clous	OUEST	60
Clous	CENTRE	100
Clous	Total	210
Vis	EST	40
Vis	OUEST	70
Vis	CENTRE	80
Vis	Total	190
Boulons	EST	90
Boulons	OUEST	120
Boulons	CENTRE	140
Boulons	Total	350
Nettoyeurs	EST	20
Nettoyeurs	OUEST	10
Nettoyeurs	CENTRE	30
Nettoyeurs	Total	60
Total	EST	200
Total	OUEST	260
Total	CENTRE	350
Total	Total	810



- Calcul des totaux: 28 accès en lecture et 8 accès en écriture
- Un SGBDR lit 200 enreg/s et en écrit environ 20/s.

	Est	Ouest	Centre	Total
Clous	50	60	100	210
Vis	40	70	80	190
Boulons	90	120	140	350
Nettoyeurs	20	10	30	60
Total	200	260	350	810



OLAP consolide entre 20 et 30000 cellules/s



Modélisation Logique – Approches OLAP

- HyperCube OLAP: représentation de l'information dans un hypercube à N dimensions



Obtenir des informations déjà agrégées selon les besoins de l'utilisateur: simplicité et rapidité d'accès

- OLAP (*On-Line Analytical Processing*): fonctionnalités qui servent à faciliter l'analyse multidimensionnelle: opérations réalisables sur l'hypercube
- Pour **implémenter le modèle multidimensionnel**, il existe plusieurs techniques:
 - ROLAP (Relational OLAP)
 - MOLAP (Multidimensional OLAP)
 - HOLAP (Hybrid OLAP)



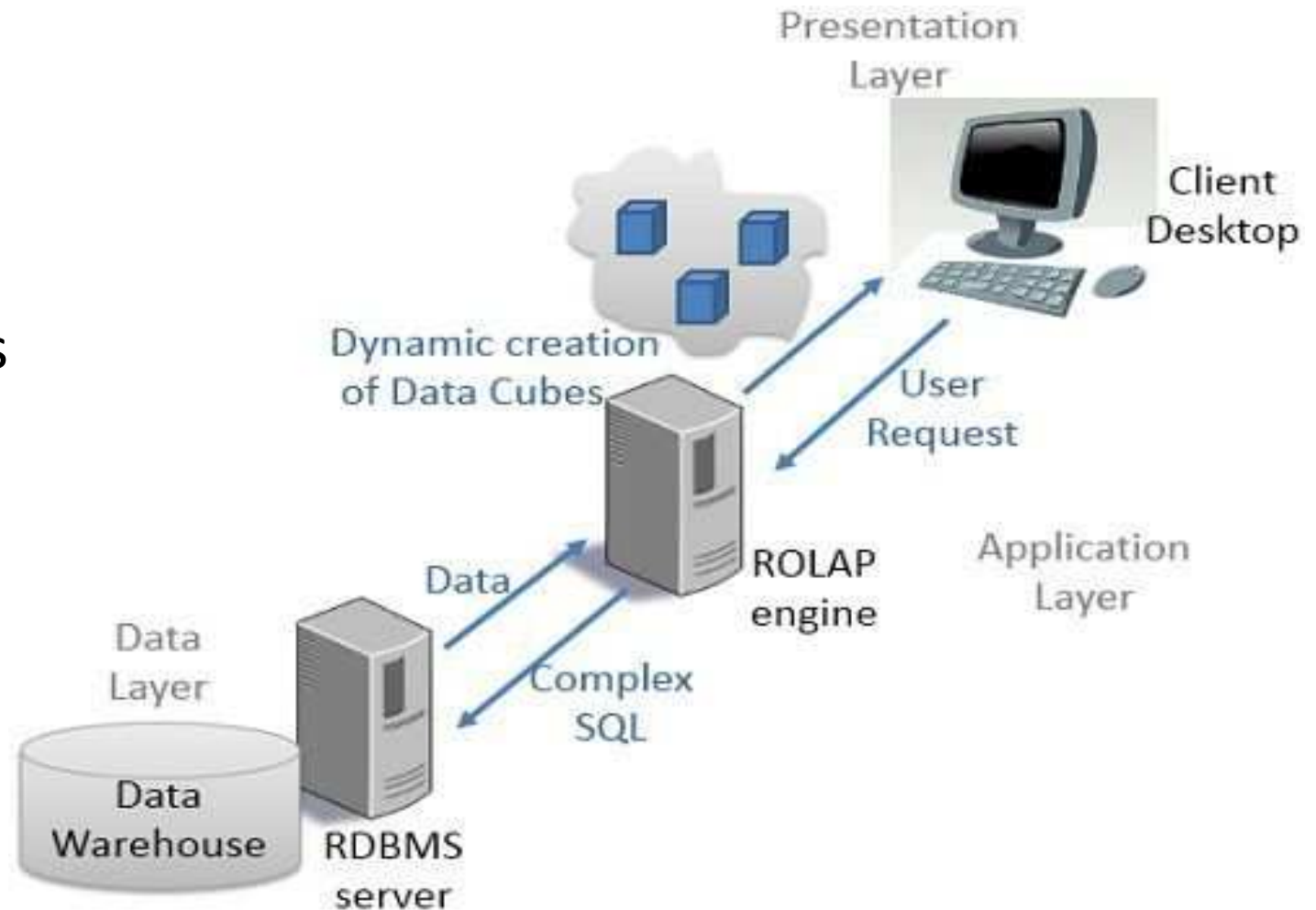
Modélisation Logique – Approche ROLAP

- Le stockage peut s'effectuer sur un SGBD relationnel classique
- Le cube est stocké selon le modèle en étoile (flocon ou constellation)
- Un moteur OLAP permet de simuler le comportement d'un SGBD multidimensionnel
 - **Traduction dynamique
Multidimensionnel/Relationnel**



Modélisation Logique – Approche ROLAP

- Peu coûteux
- Implémentation facile
- Stockage de gros volumes de données



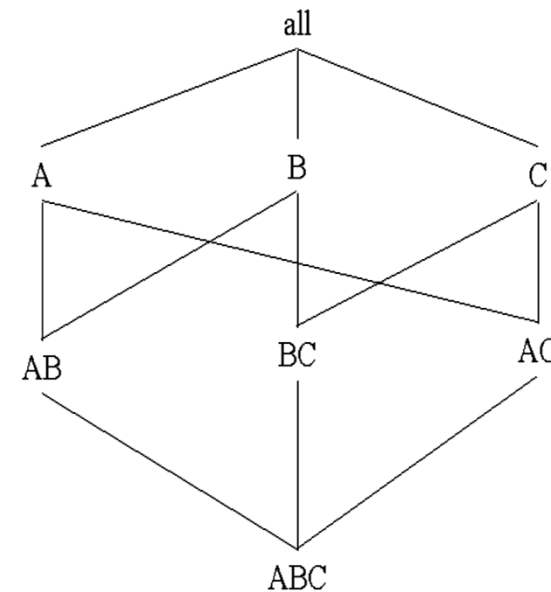
ROLAP Model

- Performance faible lors des calculs



Modélisation Logique – Approche ROLAP

- L'*UNION* de plusieurs *Group-by* donne naissance à un cube
- Opérateur cube (SQL99): généralisation N-dimensionnelle de fonctions d'agrégations simples
- Opérateurs d'agrégation : *Group-by cube*, *Group-by rollup*
- **GROUP BY CUBE (a,b,c)** est équivalent à:
 - GROUP BY (a,b,c)
 - GROUP BY (a,b)
 - GROUP BY (a,c)
 - GROUP BY (b,c)
 - GROUP BY (a)
 - GROUP BY (b)
 - GROUP BY (c)
 - Aucun GROUP BY



57



Modélisation Logique – Approche ROLAP

- Exemple:

*Select Magasin, date, Rayon, Sum(CA Ventes)
From VENTES
Group-by Cube Magasin, Date, Rayon ;*

Magasin	Date	Rayon	CA Ventes
Mag1	2/3/2020	010	3500
Mag1	10/3/2020	010	2500
Mag1	19/3/2020	010	2900
Mag1	ALL	010	8900
Mag2



Modélisation Logique – Approche MOLAP

- Les données sont stockées dans des tableaux à plusieurs dimensions, pouvant être **éparses**
- On y stocke les mesures (valeurs à observer) dans les cellules ; les données représentant les dimensions sont les coordonnées de ces valeurs :
$$C = (d1, d2, \dots dn, m1, m2, \dots mp)$$
- L'accès aux données est **direct**



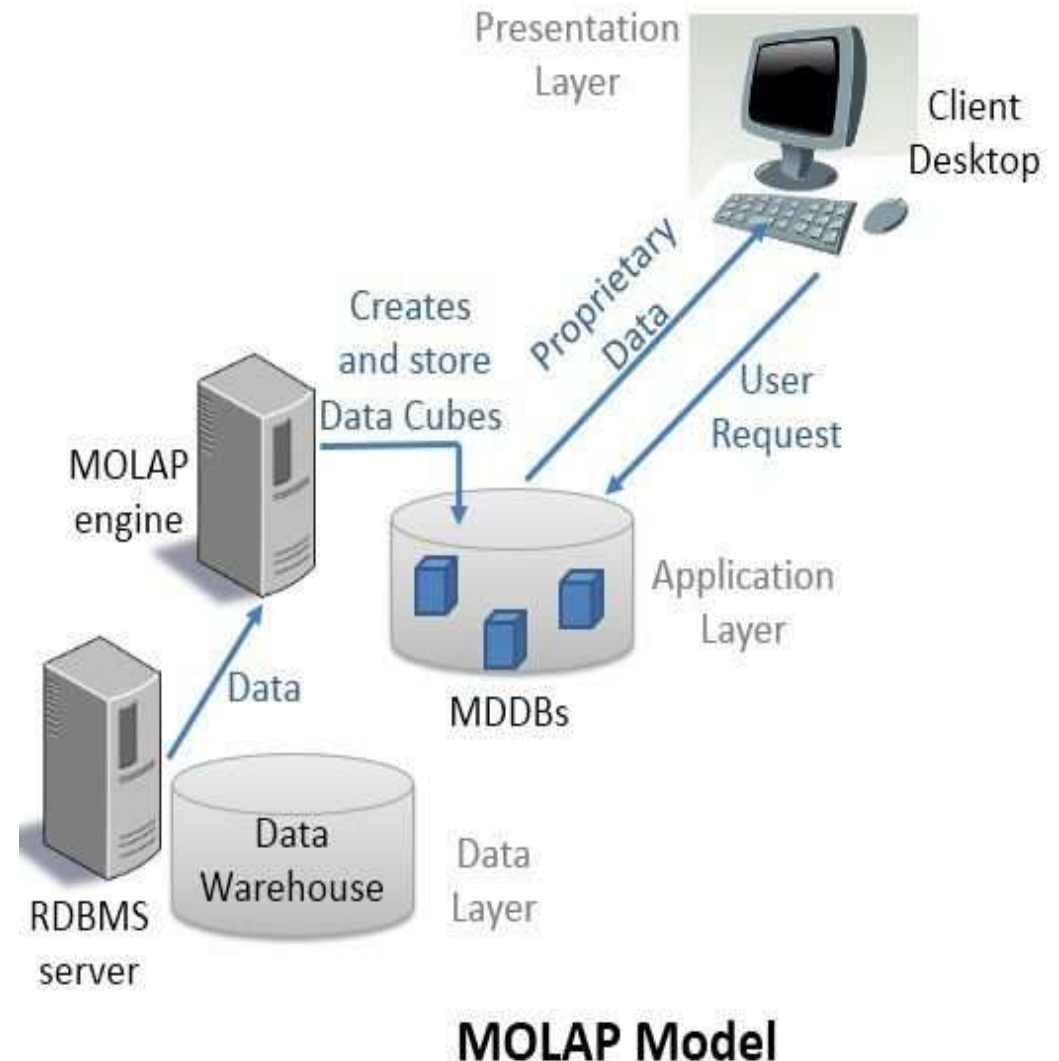
Modélisation Logique – Approche MOLAP



- Temps de calcul très rapide

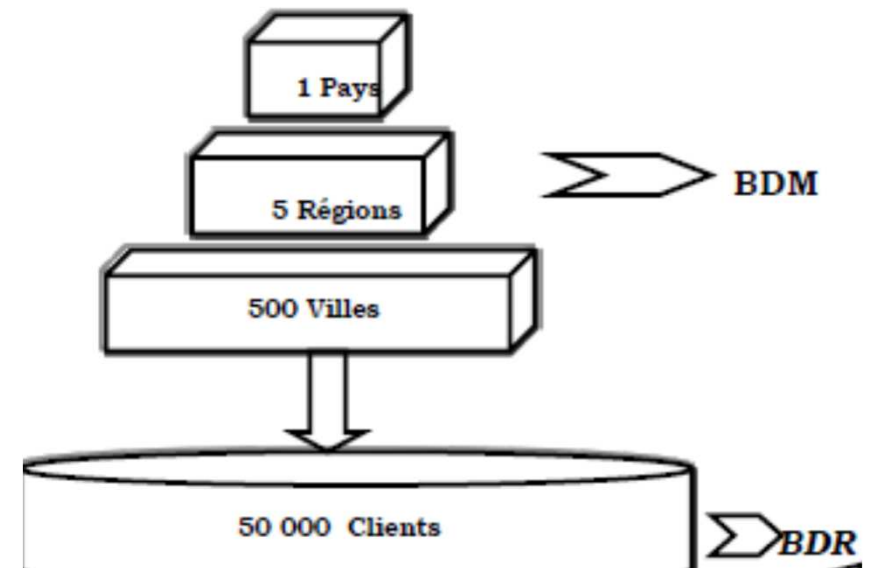


- Mise en place difficile
- Ne supporte pas de gros volumes de données
- Formats propriétaires

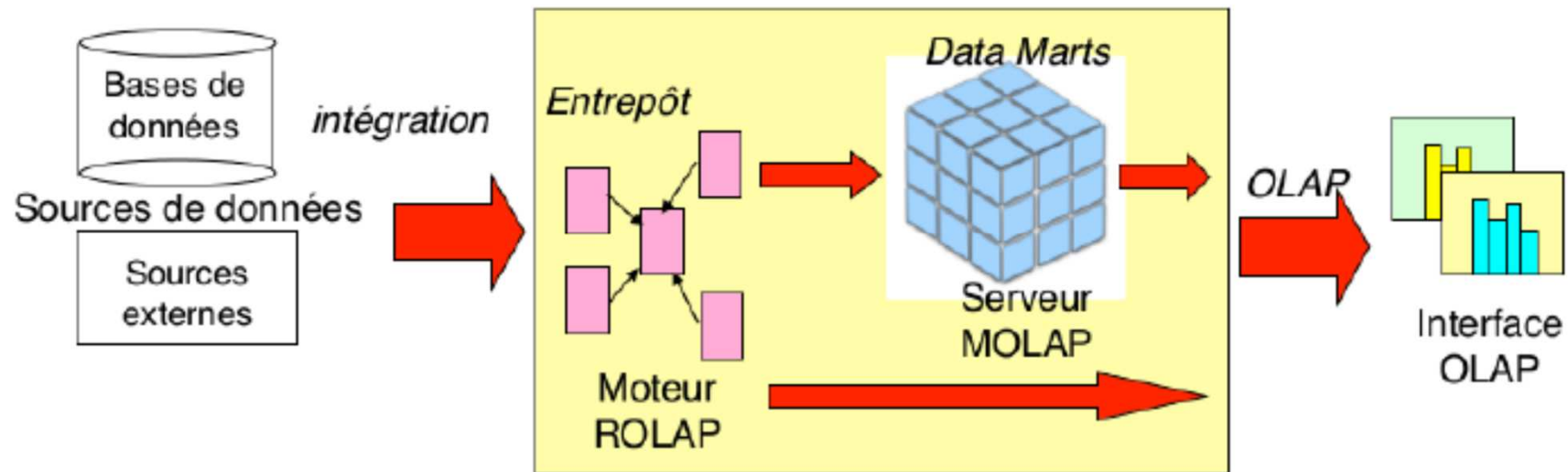


Modélisation Logique – Approche HOLAP

- Hybride:
 - Mode de stockage propriétaire pour les tables d'agrégat et les tables intermédiaires
 - Mode relationnel pour les tables avec les données fines.
- Etapes:
 1. Charger les données d'une table vers un tableau.
 2. Calculer le cube de ce tableau
 3. Stocker les résultats (données agrégées) dans un tableau



Modélisation Logique – Approche HOLAP



- Bon compromis cout/performance



- Mise en place difficile
- Moins rapide que MOLAP
- Passage à l'échelle moins facile que ROLAP



Modélisation Logique – Approches OLAP

- Solutions commerciales
 - DB2 UDB Server : ROLAP
 - Oracle : ROLAP
 - SQL Server 2000 : ROLAP
 - SAS OLAP Server : MOLAP
 - SQL Server : HOLAP
 - Oracle Express-server : MOLAP/ROLAP
 - DB2 OLAP Server : MOLAP/ROLAP



Modélisation Logique – Synthèse

Approche	ROLAP	MOLAP	HOLAP
Stockage data			
- Détaillées	Relations	Cubes	Relations
- Agrégées	Relations	Cubes	Cubes
Avantages	<ul style="list-style-type: none"> - Facile à mettre en œuvre - Peu coûteux - Evolution facile - Stockage de gros volume de données 	Temps de calcul très rapide	Compromis couts et performances
Inconvénients	Performance faible lors des calculs	<ul style="list-style-type: none"> - Mise en œuvre difficile - Formats propriétaires - Ne supporte pas de gros volumes de données 	
Outils	<ul style="list-style-type: none"> - IBM DB2 UDB - Mondrian (Pentaho) 	<ul style="list-style-type: none"> - SAS OLAP Server 	<ul style="list-style-type: none"> - Oracle Express Server - DB2 OLAP Server - SQL Server

