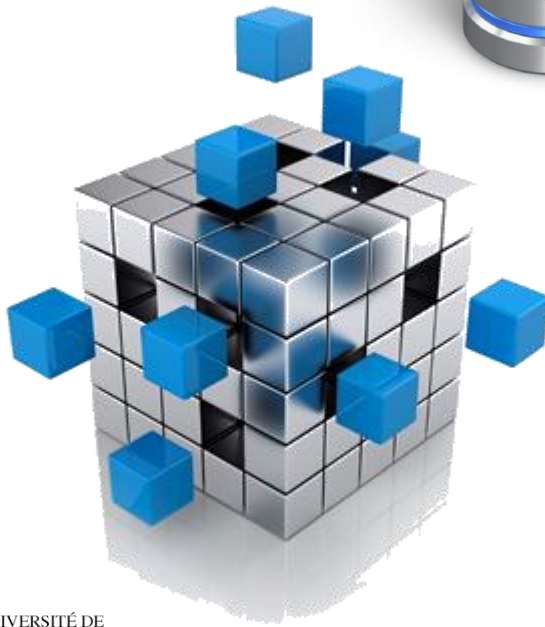


# INF 735

## Entrepôt et Forage de Données



**Bloc 7**  
Techniques de réalisation et opération  
par Robert J. Laurin

# Plan du cours – Les blocs

**(Bloc1)**

Introduction: Le besoin, concepts et définitions

**(Bloc 4)**

ETC:  
Acquisition de  
données

**(Bloc 2)**

Modélisation  
(Entrepôt)

**(Bloc 3)**

Outils de  
présentation,  
OLAP et Forage

**(Bloc 5)** Architecture et Méta données

**(Bloc 6)** Définition des besoins et gestion de projet

**(Bloc 7)** Techniques de réalisation et opération



## Suggéré:

Data Warehousing Fundamentals, A Comprehensive Guide for IT Professionals,

Paulraj Ponniah

- Chapitres 8, 18 à 20
- Appendice B
- Chapitre 16 (retour)

- Annexes:

- Deployment readiness checklist
- Desktop installation readiness checklist
- Index plan
- Security checklist
- Technical architecture plan
- Gartner identifies the `Fatal Flaws` of business intelligence

- L'entrepôt se veut aussi WEB
- Passer du model logique au physique = rigueur
- Infrastructure Informationnelle  $\neq$  opérationnelle
- Intégrer la gestion de l'entrepôt à l'opération

---

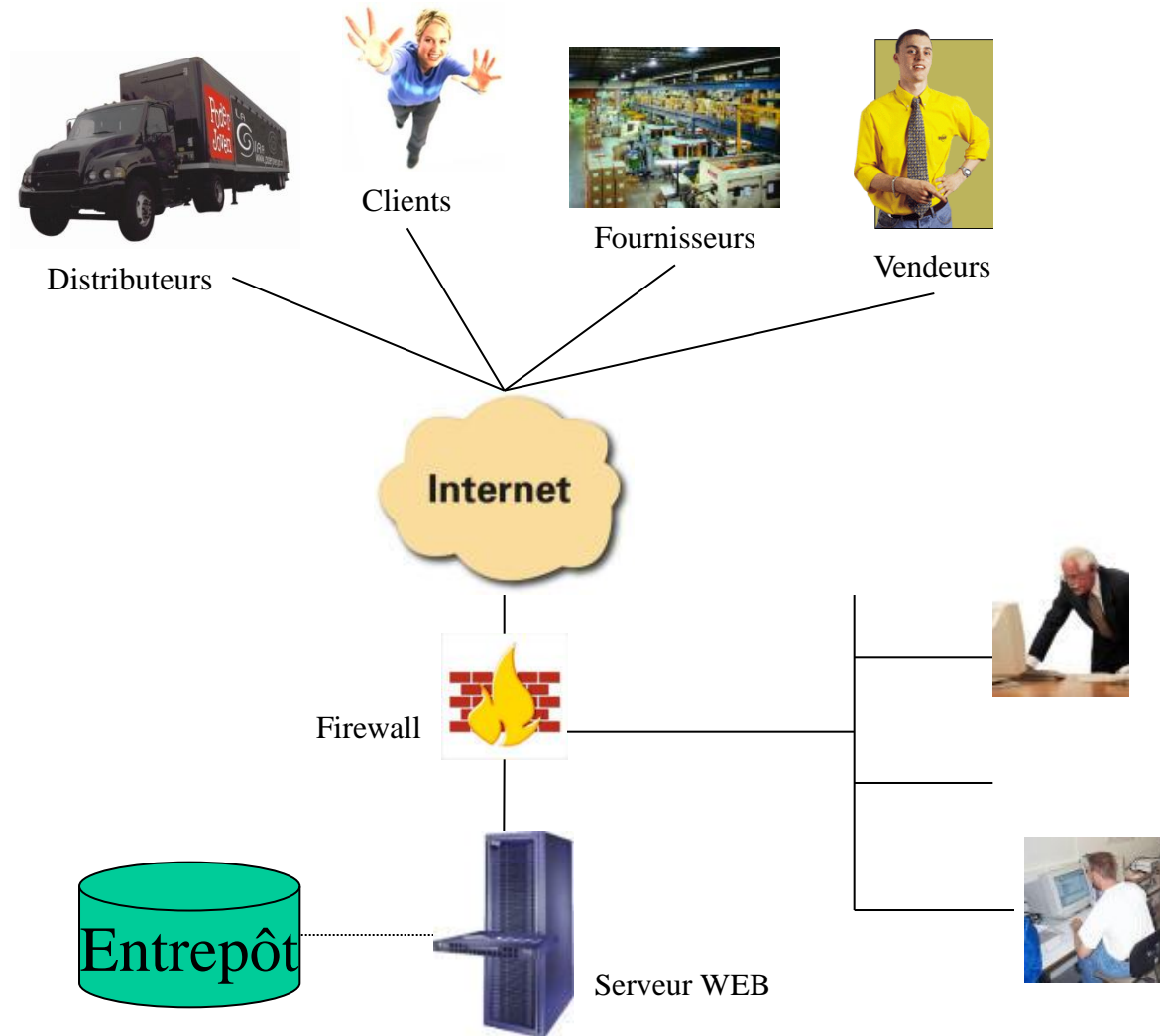
# DATA WEBHOUSE

# L'entrepôt et le Web/Intranet

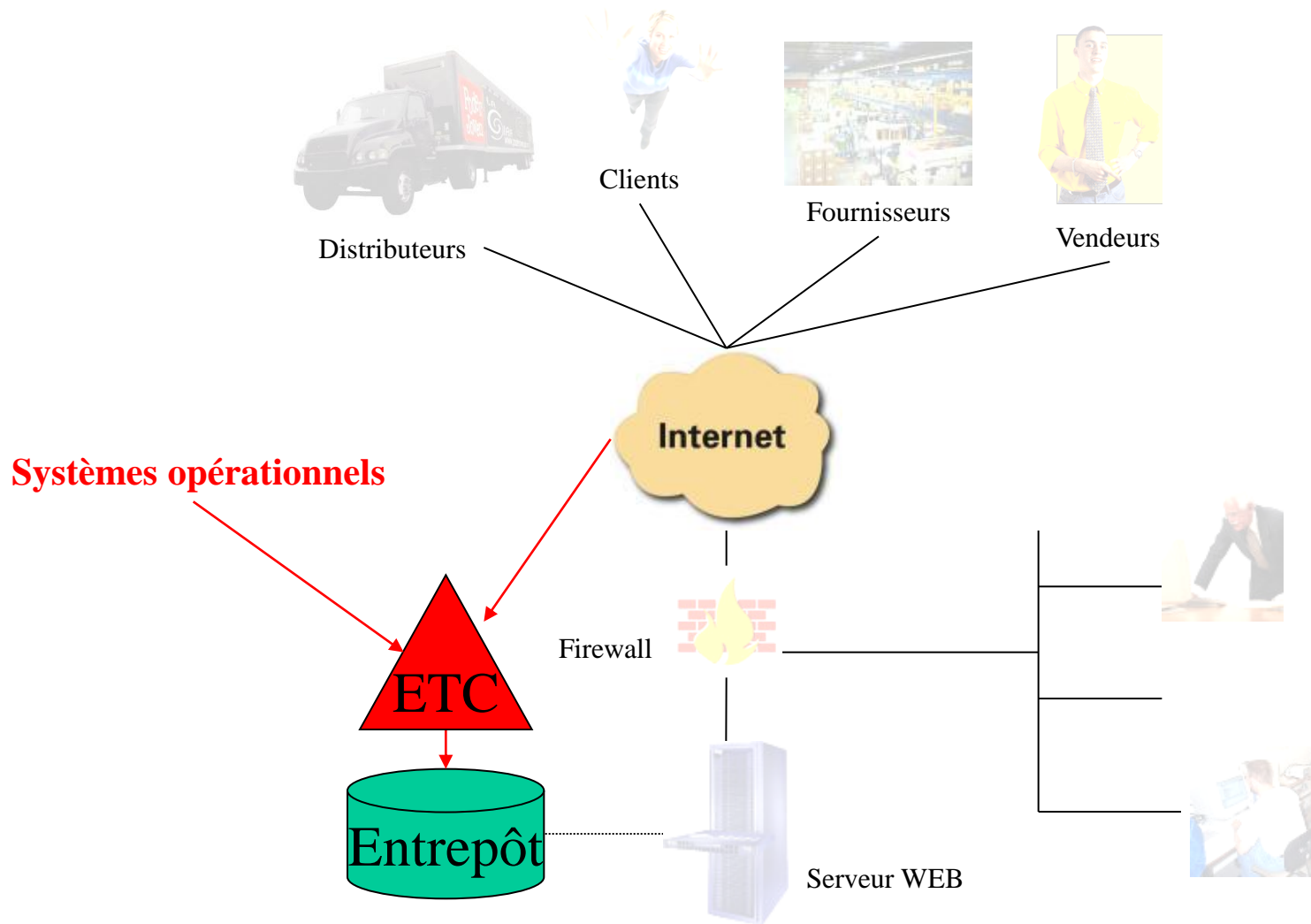
---

- Depuis 1990 – Internet et Intranet sont des incontournables
  - « Data WEBhouse », Dr. Ralph Kimball, 1999
- Accès à un complément d'information pour l'entrepôt (vers l'intérieur)
- Publication de l'information stratégique au monde (<https>)
- Facilité de déploiement ← interface universelle
- Permet de livrer de l'information stratégique à moindre coût.
- Une seule source pour tous les outils
- Diminue les besoins en formation

# L'entrepôt et le Web/Intranet - Fournisseur

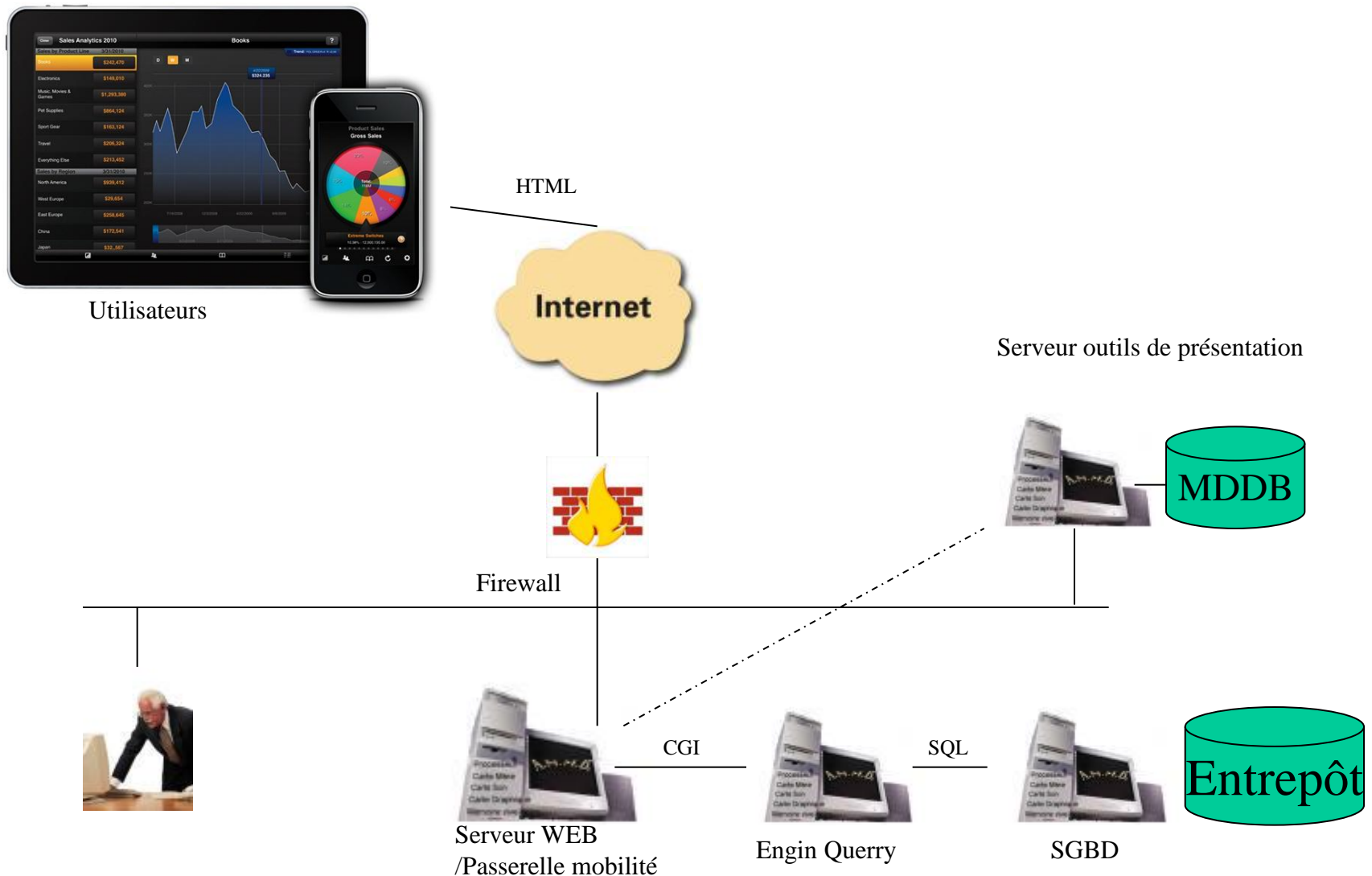


# L'entrepôt et le Web/Intranet - **Consommateur**





# Entrepôts – Éléments techniques



---

# PASSER DU MODÈLE LOGIQUE AU PHYSIQUE

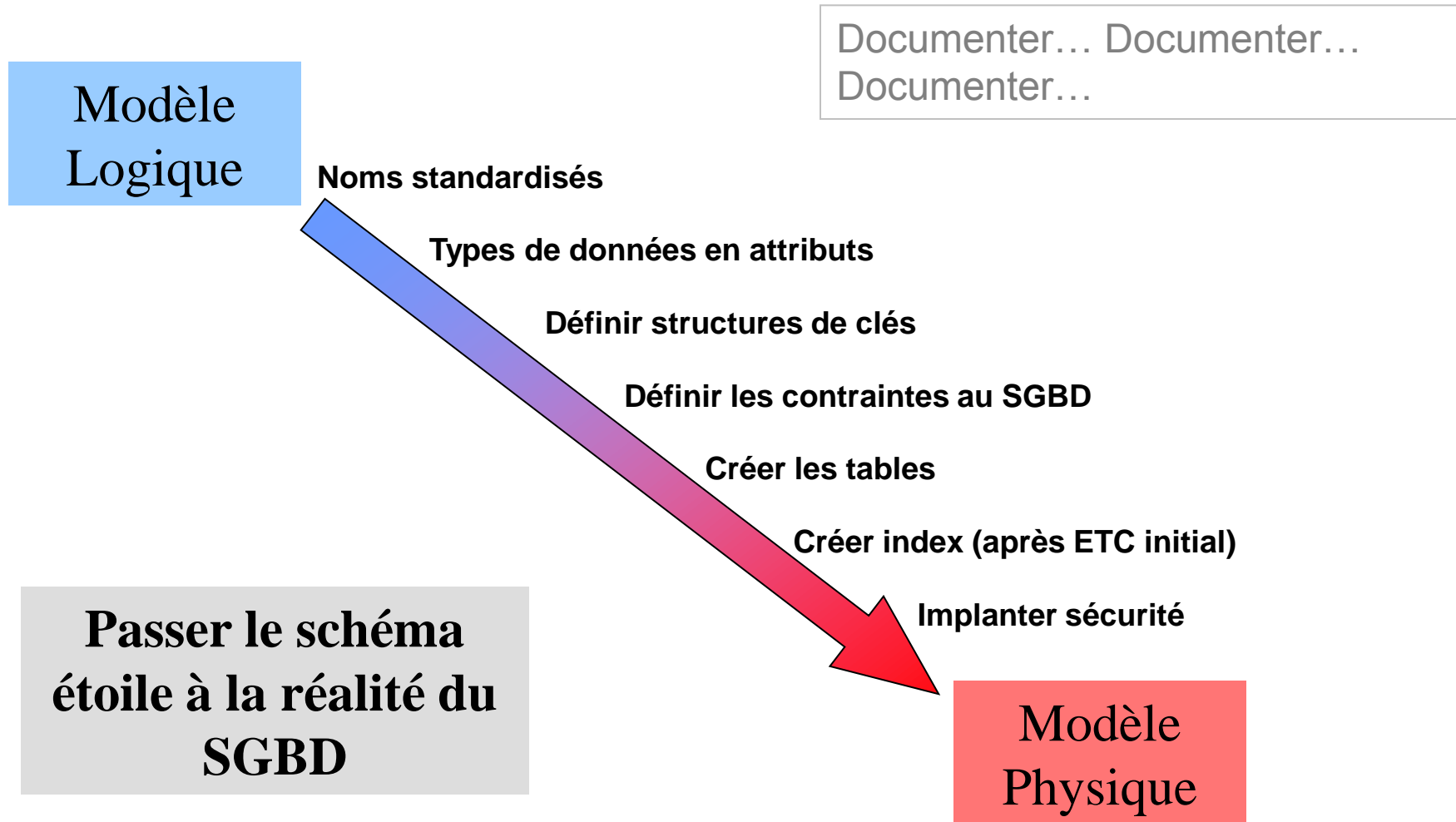
# Passer du modèle logique au physique - étapes

- Développer les standards:
  - Nomenclatures
  - Procédures de travail
  - Formulaire
  - Suivi de projet
  - Méthode de documentation
- Analyser et créer les agrégats
  - Analyser les besoins et ajuster le design
- Déterminer le schéma de partitionnement
- Établir les option de regroupement
- Préparer la stratégie d'indexation
- Compléter le model physique

```
Create Table Dim_client  
(Client_key Int not null,  
Nom ...);
```

```
Create Table Fact_sales  
(...);
```

# Passer du modèle logique au physique - étapes



# Modèle physique - Performance

- Index sur champs
  - Attention au Chargement
  - Attention aux index incomplets
  - Attention à avoir trop d'index (AU CHARGEMENT)
  - Sélection sur clés primaires et étrangères d'abord.
- Index sur table de faits:
  - B-tree sur l'ensemble des clés étrangère ← clé
  - Placer les clés dans l'ordre descendant des probabilités de requête
  - Créer des index sur les éléments les plus sollicités
  - Possiblement indexer les faits si font régulièrement parti des contraintes
- Index sur table dimension:
  - B-tree sur clé
  - Bitmap sur options populaires (colonnes)
  - Indexer les groupes de colonnes souvent sollicitées (ordre de probabilité descendant)
  - Index sur éléments possibles de joint

+ matériel  
+ Partitionnement  
+ Cluster  
+ Traitement Parallèle  
+ Agrégation

# Modèle physique - Performance

---

Les Indexes prennent souvent plus d'espace que les données dans un entrepôt!

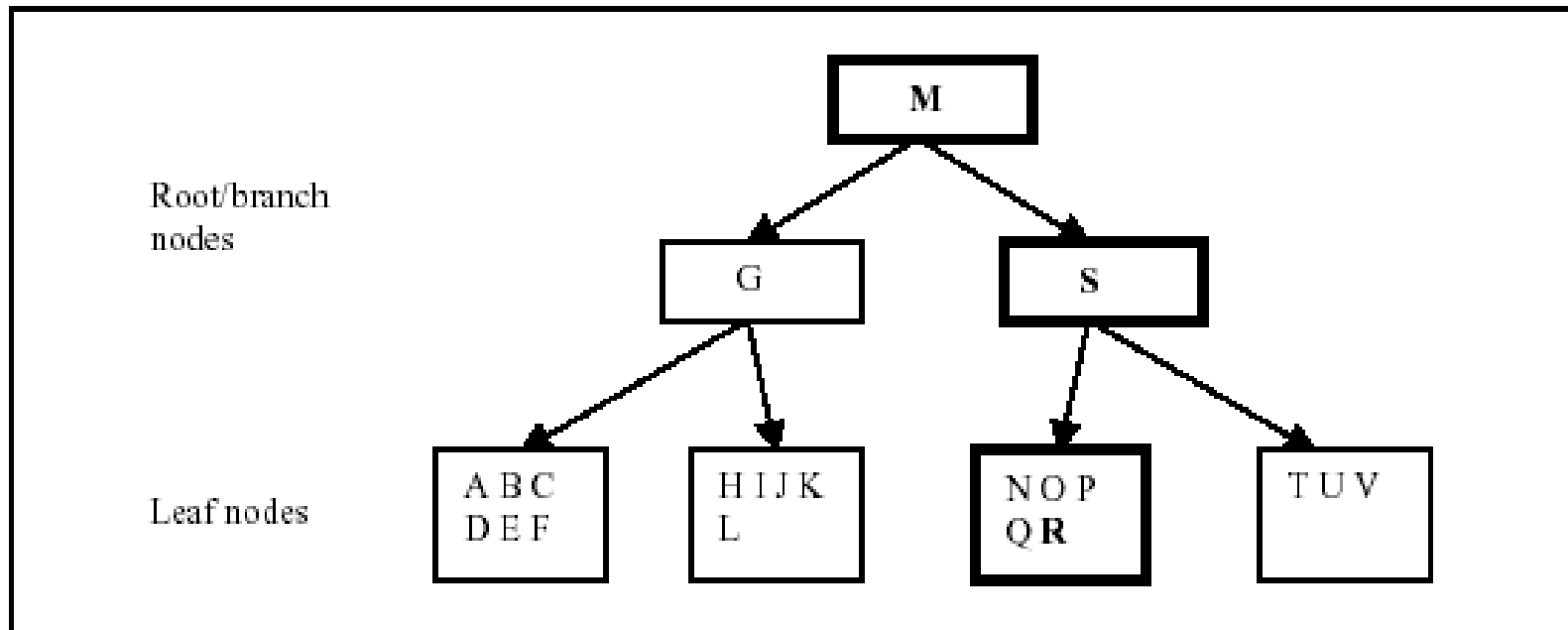
- Lecture seulement → Aucune préoccupation quant au nombre d'écritures requises
  - + Sauf lors d'alimentation où il est préférable d'arrêter l'indexation
- Optimiser pour la performance de la recherche = + d'indexes

# Index – Bit-mapped

---

Valeurs du domaine / Nom du client	Célibataire	Marrié	Divorcé	Veuf	Pattern
Grand Talent	1	0	0	0	<b>1000</b>
Petit Trot	0	0	1	0	<b>0010</b>
Gros Louis	0	0	0	1	<b>0001</b>
Grand Galop	0	1	0	0	<b>0100</b>

La recherche par statut marital peut se faire sur l’index Bit-map.



*Fig. 1a : A simplistic B-tree implementation (height : 2 and order : 2) [Branch/leaf nodes read to locate "R" are displayed in bold]*

## Sous types de B-Tree:

- Partitioned indexes
- Reverse-key indexes
- Index-organized tables (IOTs)
- Cluster indexes

## The Art of Indexing in Oracle8 i

Venkat S. Devraj

*IMMEDIANT (formerly Raymond James Consulting)*



---

# INFRASTRUCTURE

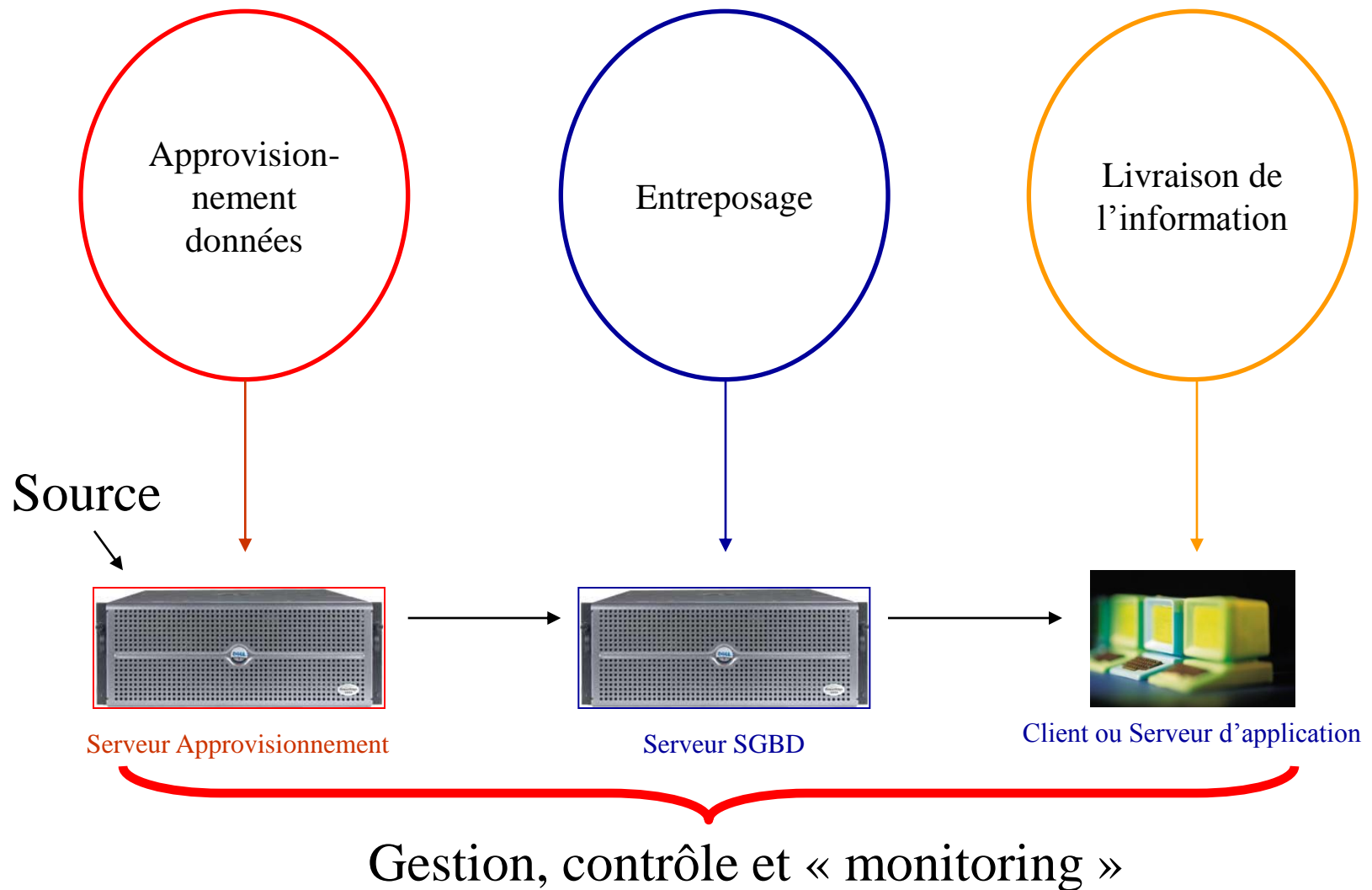
# Modèle physique - Considérations

- Prévoir la croissance
  - CPU, mémoire , Disques...
- Gestion de l'espace disque
  - Raid 5 ok → lecture surtout
  - Plusieurs « petits » disques
- Outils d'administration et monitoring
- Chercher la flexibilité

Infrastructure  
séparée

→ Séparation du monde opérationnel et informationnel,  
même en infrastructure

- Hardware et Software
  - Outils acquisition
  - Outils d'approvisionnement en information
  - Outils de forage, rapports, EIS, OLAP, et autres de livraison de l'info.
  - SGBD et Serveurs/OS/NOS
  - Réseautique, Télécoms.
  - Outils de protection, Protections électriques, antivirales...
  - Gestion de l'environnement et transactions



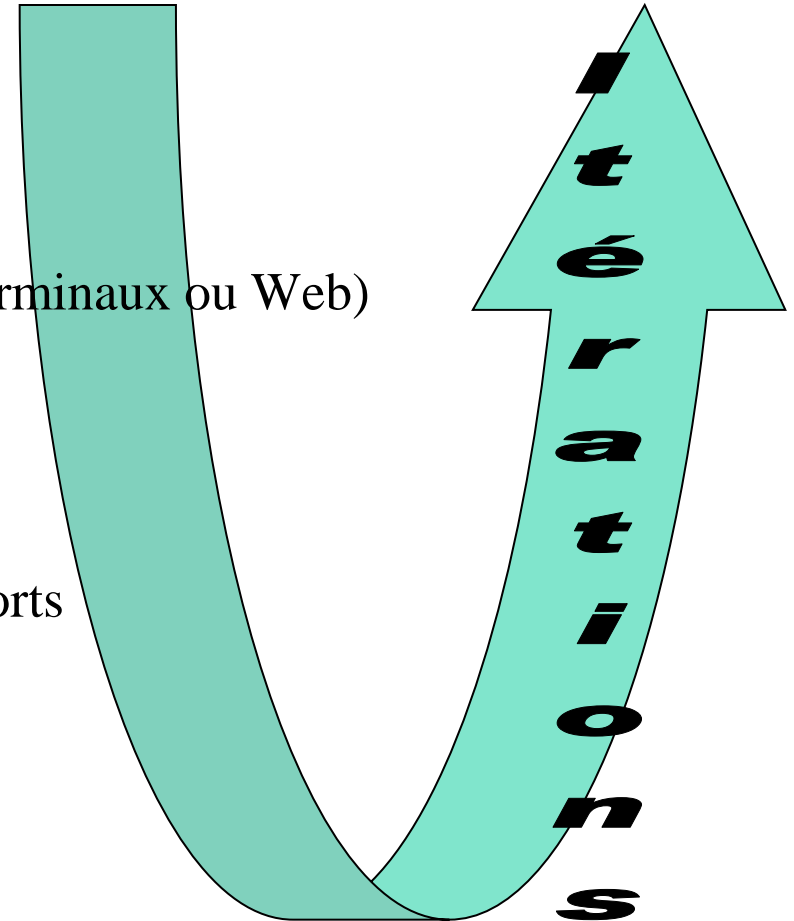
- Choix technologique teinté par:
  - l'architecture/infrastructure déjà en place
  - Le besoin en traitement et stockage
  - La plate-forme serveur et la stratégie de parallélisme
  - Les options du jour et le vendeur du mois... (moins scientifique)
- SGBD d'entrepôt :
  - Oracle
  - DB2 (IBM)
  - Microsoft SQL
  - MySQL
  - Autres?

---

# DÉPLOIEMENT ET SUPPORT

# Déploiement – Liste d'activités

- ETC initial
- Approbation utilisateur
- Préparer les postes de travail (ou Terminaux ou Web)
- Formation des utilisateurs
- Faire les premières requêtes et rapports
- Support aux utilisateurs
- Déployer progressivement



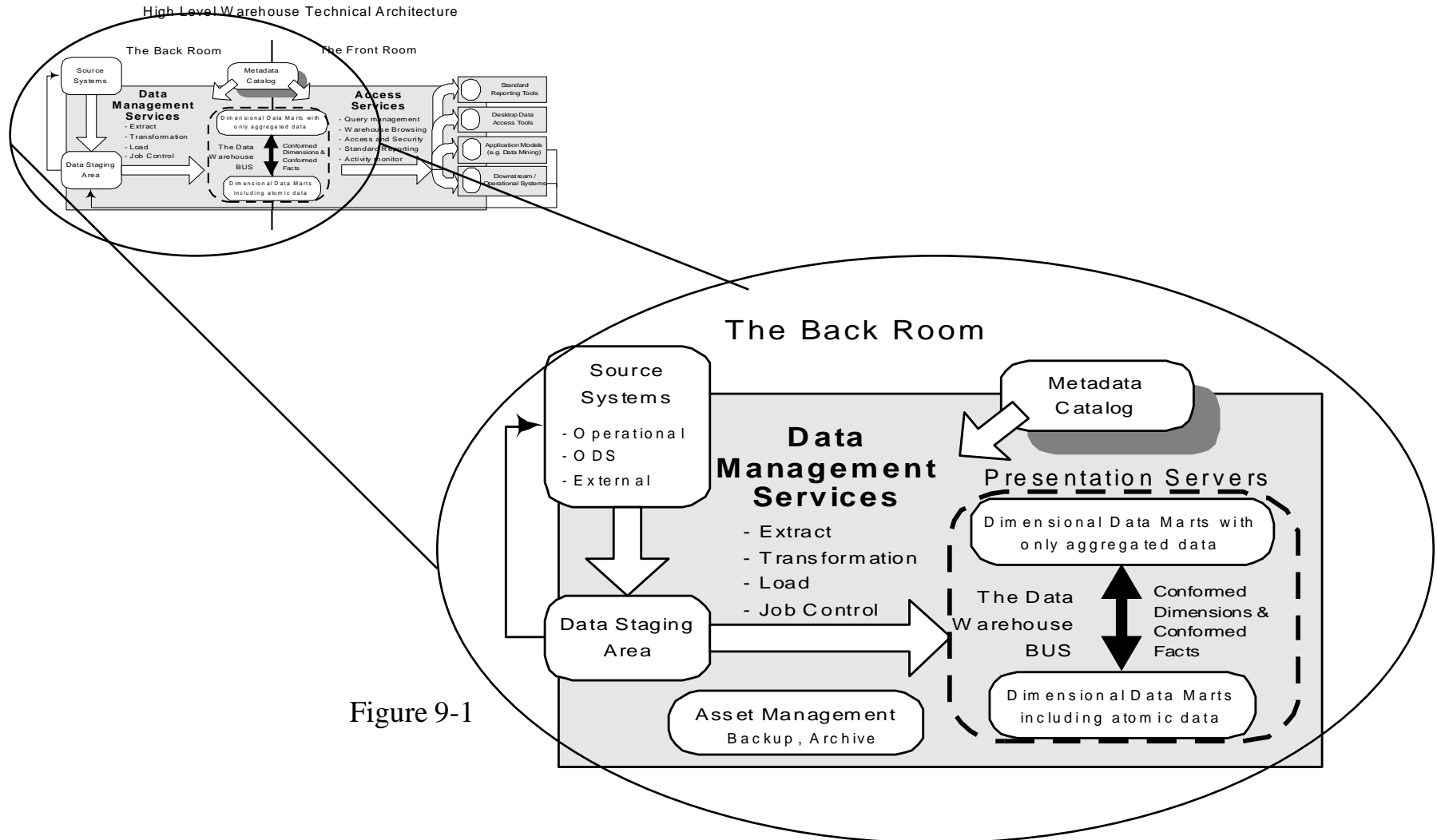
- Modèle de données ! ! !
- ETC – Fréquences des mises à jour / sujet
- Outils disponible
- La méta-donnée
- Formats standardisés (dates, etc...)
- Conciliation ou non avec les systèmes opérationnels

Un utilisateur ne devrait pas avoir accès sans être préalablement formé.  
Changer le programme d'accueil ...



- Mesurer et adapter...
- Intégrer au support déjà fourni aux utilisateurs
- Gestion de la technologie intégré aux autres
  - Plate-forme
  - Backup
  - Sécurité
  - Réseau
  - Télécoms
  - Upgrades
  - Alertes ?
- Monitoring du ETL

# Gestion post-implantation



# Entrepôt dysfonctionnel?

---

- Le promoteur est:
  - Toujours en poste ?
  - Utilisateur de l'information et l'entrepôt ?
  - Pas un suiveur de mode pour bien paraître ?
- L'équipe de projet comprends les besoins et objectifs des utilisateurs; et les utilisateurs ont confiance que l'équipe comprends...?
- Les utilisateurs peuvent comprendre le modèle dimensionnel, qui a été monté de façon à répondre aux besoins et non à des requis T.I.?
- Le besoin analytique des utilisateurs est comblé par des outils adéquats?
- Les données sont justes et réconciliables/explicables ?
- Les données sont rafraîchies périodiquement selon l'horaire prévu et publié?
- Les utilisateurs ont été adéquatement formés ?
- Les utilisateurs comprennent les données et leur provenance?
- Chacun sait qui appeler quand il a des problèmes ?

# Gartner identifies the `Fatal Flaws` of business intelligence and advises organisations on how to avoid them ©1994-2005

---

Gartner survey of 1300 CIOs in more than 30 countries around the world has revealed that CIOs plan to increase spending on BI by an average of 6 percent this year. As companies move away from cost cutting initiatives, CIOs believe strategic use of business intelligence, combined with a focus on improving business processes, will be most significant in delivering IT's contribution to business growth in 2005 through 2008.

Gartner's leading BI analysts highlighted seven major flaws inherent in BI and outlined Gartner's best practice for avoiding them.

Flaw 1 - "If we build it, they will come"

Flaw 2 - "Managers need to negotiate the numbers"

Flaw 3 - "Data quality problem we don't have one"

Flaw 4 - "Our enterprise applications vendor will deliver the best solution"

Flaw 5 - "Darwin was right - BI projects need to evolve"

Flaw 6 - "We can outsource the whole thing"

Flaw 7 - "Just give me a dashboard!"

Gartner provided three key recommendations that follow five years of study drawn from thousands of customer interactions and codified in Gartner's BI Framework.

1- Make sure you have senior level business sponsorship for BI

2- Have a unified BI infrastructure

3- Leverage existent wisdom and evolve your BI initiatives.

Source: Gartner 2005

---

# CONCLUSION

- Outil informationnel efficace
- Qui dégage les systèmes opérationnels
- Qui amène une version unique de la vérité
- Avec un ensemble de techniques et d'outils différents
- Offrant une vision plus « corporative » ou harmonieuse des données
- Amène donc une gestion de projet plus « politique »
- Axée sur l'intégration des données et des processus d'affaires
- Qui soulève tous les problèmes de qualité de données,
- de processus mal arrimés,
- des priorités d'entreprise conflictuelles,
- et même de la qualité des gestionnaires