

# **BASES DE DONNÉES**

## ***ENTREPÔTS DE DONNÉES***

### **Une introduction**

BD300  
v100b

2016-10-15

Christina KHNAISSER et Luc LAVOIE  
Département d'informatique  
Faculté des sciences



Christina.Khnaisser@usherbrooke.ca  
Luc.Lavoie@usherbrooke.ca  
<http://info.usherbrooke.ca/llavoie>

# PLAN

- Problématique
- Le concept d'entrepôt de données
- Les architectures
- Les modèles de schéma
- Les visions
- La question du temps
- La question des origines
- Une vision évolutive et la renormalisation
- Quelques questions ouvertes



# PROBLÉMATIQUE

- Une BD c'est bien, deux c'est mieux, trois c'est trop?
  - de l'ordinateur central
  - en passant par l'architecture client-serveur,
  - le dépôt départemental et
  - les BD personnelles
  - on arrive au chaos!
- Nécessité d'avoir une image complète cohérente
  - Comment gérer la multiplicité ?
  - Comment unifier l'hétérogénéité ?

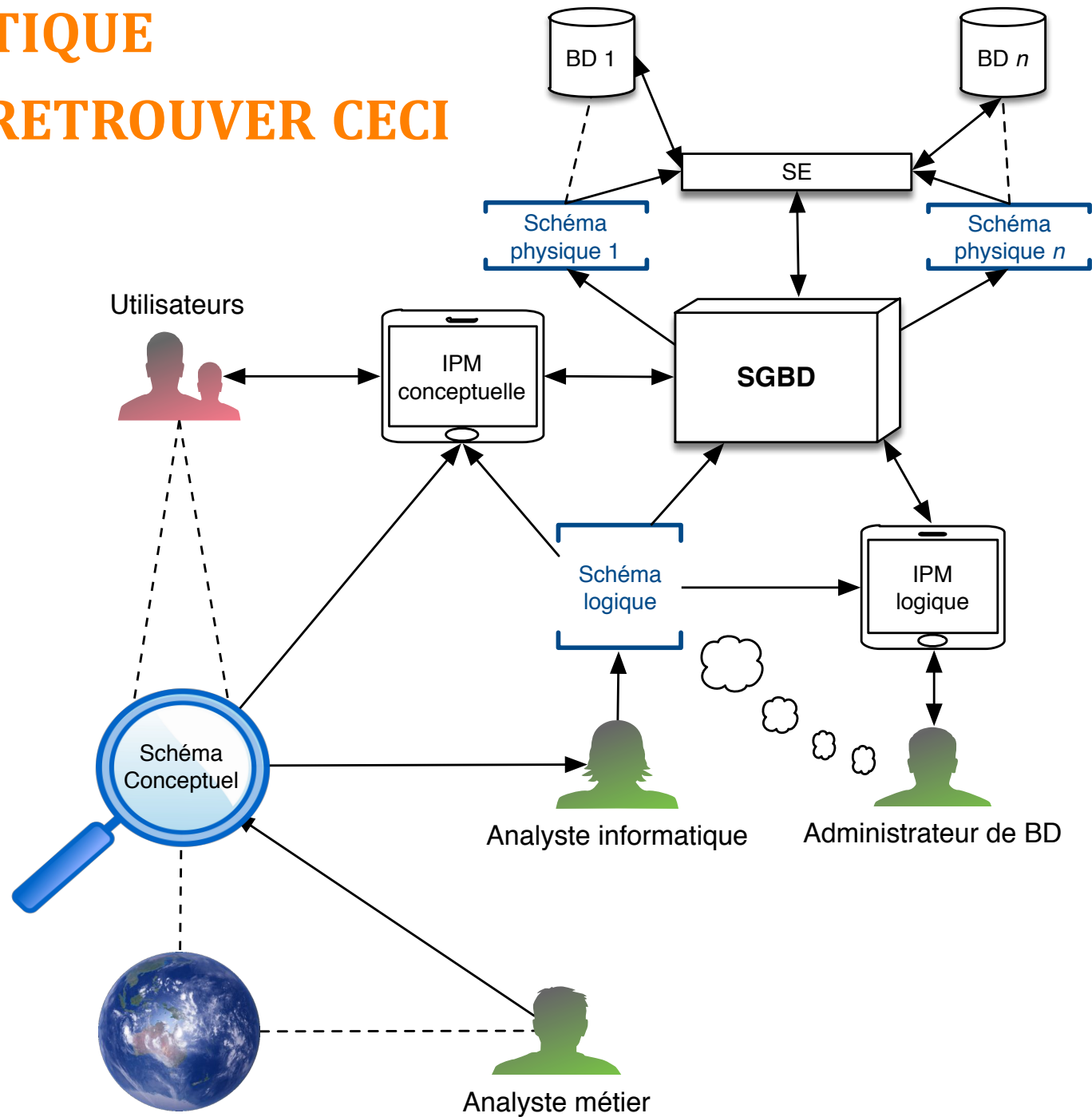
# PROBLÉMATIQUE

## EXPLOITATION ET ANALYSE, DES BESOINS DIFFÉRENTS ?

	<i>Exploitation</i>	<i>Analyse</i>
Objectif	Soutenir les processus	Mesurer les processus
Fonctions	CRUD	R ( <i>ucd</i> )
Cible d'optimisation	U (r)	R
Portée	Transaction	Lot de transactions
Nature des requêtes	Prédéfinie et stable (souvent plus de 90%)	Prédéfinie et stable (souvent moins de 50%)
Temporalité	Courante	Historique (surtout passé)
Principes de conception	Consensus : Normalisation (3FN, FNBC, 5FN)	Débat : Dimensionnalité, 6FN
Appellations courantes	<b>Système transactionnel</b> On Line Transaction System (OLTP) Système source	Système analytique <b>Entrepôt de données</b> Data Mart Data Warehouse

# PROBLÉMATIQUE

## COMMENT RETROUVER CECI



# ENTREPÔT DE DONNÉES

- Comment « fédérer » les « visions » ?
  - Agrégation des « faits »
    - par les processus « producteurs » de données,
    - donc selon l'axe du temps.
  - Documentation des « dimensions »
    - par les entités liées.
  - Découpage des sources par producteur.
  - Liaison des sources par entité.

# ENTREPÔT DE DONNÉES

## ARCHITECTURES

- Data Mart (DM)
  - la pratique originale « ad hoc »
- Information Factory (IF)
  - la proposition originale de Inmon
- Data Warehouse (DW)
  - la proposition subséquente de Kimball

# ENTREPÔTS DE DONNÉES

## MODÈLES DE SCHÉMA CLASSIQUES

- Cube
  - décomposition par entités indépendantes
  - structuration orthogonale
- Étoile
  - décomposition par entités indépendantes
  - structuration par processus producteur
- Flocon
  - décomposition par entités indépendantes
  - hiérarchisation des entités
  - structuration par processus producteur



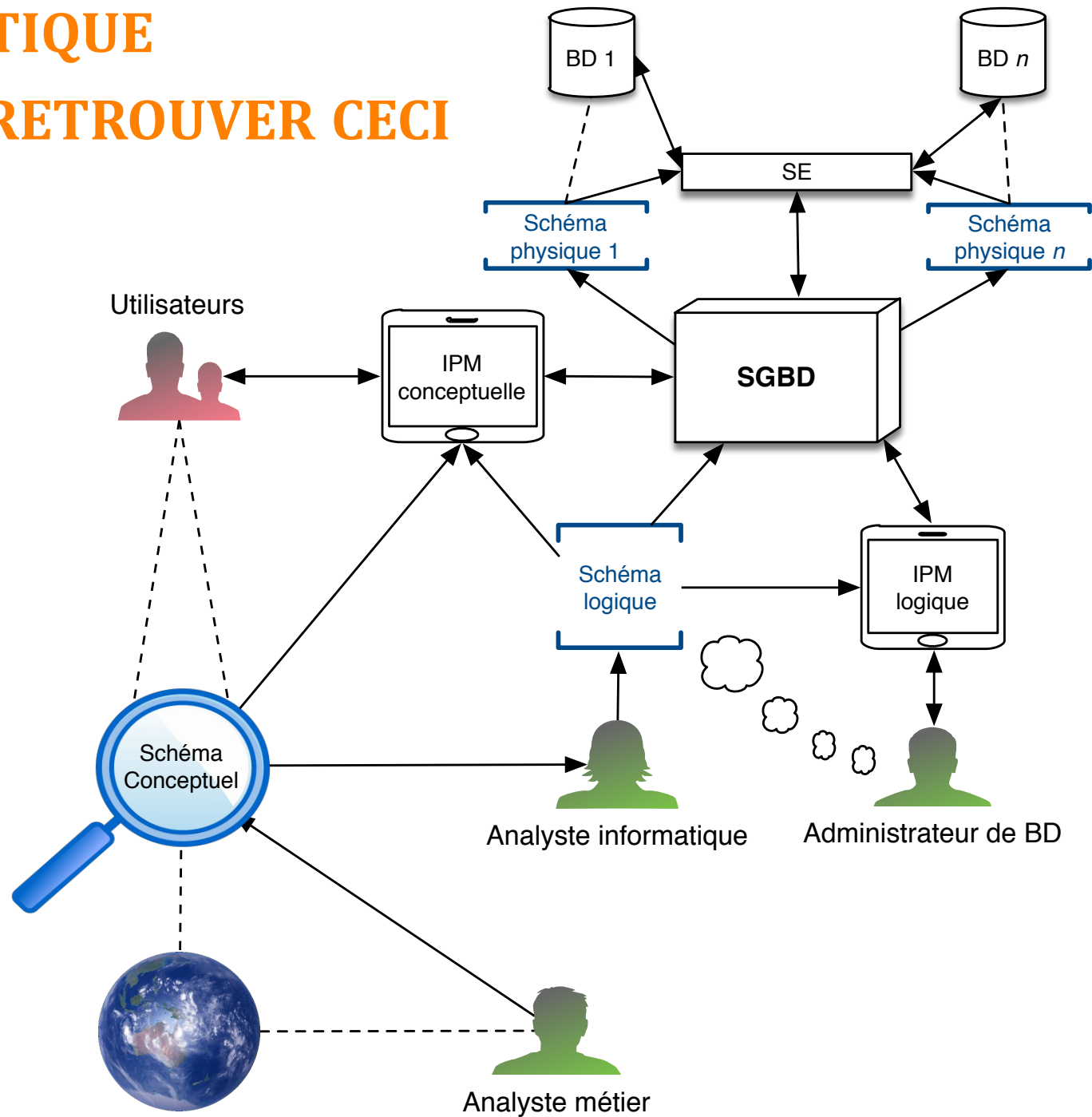
# ENTREPÔTS DE DONNÉES

## VISIONS

- Vision classique
  - temporellement estampillée
  - incrémentale pure
  - optimisation par la dénormalisation
    - quelle normalisation ?
    - quelle optimisation ?
- Vision contemporaine
  - résolument bitemporelle
  - ne peut être seulement incrémentale
  - optimisation par la normalisation
    - quelle méthode d'accès ?
    - quelle gestion mémoire ?

# PROBLÉMATIQUE

## COMMENT RETROUVER CECI



# LA QUESTION DU TEMPS

- De la mesure du temps
- Temps et calendrier
- Relativité du temps
- Modèle simple
  - discret (chronon)
  - fini (alpha, omega)
  - intervalles
  - opérateurs de Allen
- Types de coordination
  - astronomique
  - entre points relativement stables entre eux

# LA MESURE DU TEMPS

- Voir Manthey

# TEMPS ET CALENDRIER

- Voir Manthey

# RELATIVITÉ DU TEMPS

- Voir Manthey

# UN MODÈLE

Modèle simple

discret (chronon)

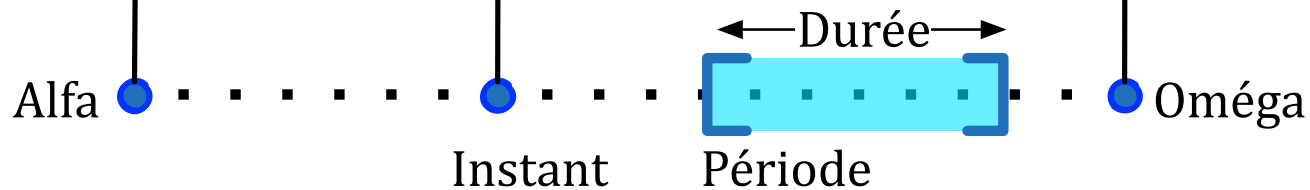
fini (alpha, omega)

intervalles

opérateurs de Allen

Perception du temps (continue)

$-\infty$   $+\infty$



Modèle discret du temps

# OPÉRATEURS DE ALLEN

Opération	gauche	droite	résultat	Description	Définition [b:e]	Allen
intervalle(b,e)	point	point	intervalle		$\{x \in T \mid b \leq x \leq e\}$	[p-:p+)
début		intervalle	point		p.b	p-
fin		intervalle	point		p.e	p+
prédécesseur		point	point		p.b-1	?
successeur		point	point		p.e+1	?
dénombrément		intervalle	entier		nombre d'éléments	?
alpha		Type point	point		La plus petite valeur du type T	first
oméga		Type point	point		La plus grande valeur du type T	last
appartenance	point	intervalle	booléen		$x \in p$	$x \in p$
appartenance <sup>-1</sup>	intervalle	point	booléen		$p \ni x$	?
antériorité stricte	intervalle	intervalle	booléen	p :  --- ○○○○○ q : ○○○○○ ---	$p.e+1 < q.b$	before
adjacence antérieure	intervalle	intervalle	booléen	p : ○ --- ○○○○○ q : ○○○○○ ---	$p.e+1 = q.b$	meets
chevauchement antérieur strict	intervalle	intervalle	booléen	p : ○○ --- ○○○○○ q : ○○○○ --- ○	$p.b < q.b \wedge p.e \geq q.b \wedge p.e < q.e$	overlaps
commencement strict	intervalle	intervalle	booléen	p : ○○○ --- ○○○○○ q : ○○○ --- ○○○	$q.b = p.b \wedge p.e < q.e$	starts
inclusion bi-stricte	intervalle	intervalle	booléen	p : ○○○○ --- ○○○○○ q : ○○○ --- ○○○	$q.b < p.b \wedge p.e < q.e$	during
achèvement strict	intervalle	intervalle	booléen	p : ○○○○ --- ○○○ q : ○○○ --- ○○○	$q.b < p.b \wedge p.e = q.e$	finishes
égalité	intervalle	intervalle	booléen	p : ○○○ --- ○○○ q : ○○○ --- ○○○	$q.b = p.b \wedge p.e = q.e$	equals



# LA COORDINATION ET LA QUESTION DES ORIGINES

- Voir Manthey

# QUELQUES QUESTIONS OUVERTES

- Traitement de l'incertitude des mesures
  - marges d'erreur
- Traitement de l'incertitude des faits
  - monde ouvert vs monde fermé
- L'alimentation : ETL ou ELT
  - origine : ELT
  - règle de pratique : ETL
  - retour à la raison : ELT

# VISION ÉVOLUTIVE

- Incontournable évolutivité
- Inadéquation de la vision incrémentale pure
  - temporalisation
  - correction
  - harmonisation
- L'entrepôt de données, une base de données comme les autres ?

# PRINCIPALES RÉFÉRENCES



- ADAMSON, C. 2010.  
*The complete reference star schema.*  
McGraw-Hill, New York, NY, USA.
- ADAMSON, C. 2008-2015.  
<http://blog.oakonsoftware.com>
- INMON, W.H. 2005.  
*Building the data warehouse.*  
John Wiley, Indianapolis, IN, USA.
- JIANG, B. 2015.  
*Constructing Data Wharehouses with Metadriven Generic Operators, and more.*  
2nd ed., Createspace.
- KIMBALL, R. 2013.  
*The data warehouse toolkit: the definitive guide to dimensional modeling.*  
John Wiley, Indianapolis, IN, USA.
- MANTHEY, R. 2015.  
<http://www.iai.uni-bonn.de/III/lehre/vorlesungen/TemporalIS/SS15/>