

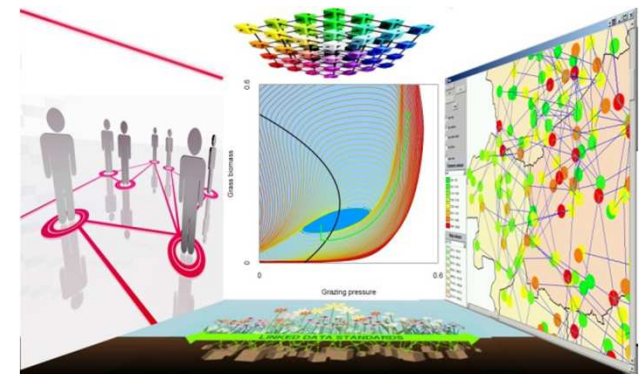
Gestion et valorisation des données capteur à l'échelle de l'exploitation dans le contexte du Big Data

Sandro Bimonte
Irstea, TSFC, Clermont Ferrand

Sandro.bimonte@irstea.fr



- Thème de Recherche MOTIVE





Plan

Panorama des technologies de GeoBusinnes Intelligence

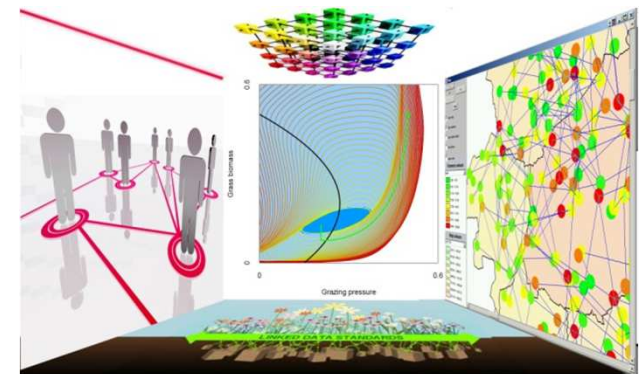
Données capteur à l'échelle de l'exploitation

Intégration des données capteur dans les systèmes GeoBI

Et le Big Data?

Conclusion

Un panorama des technologies de GeoBusinnes Intelligence



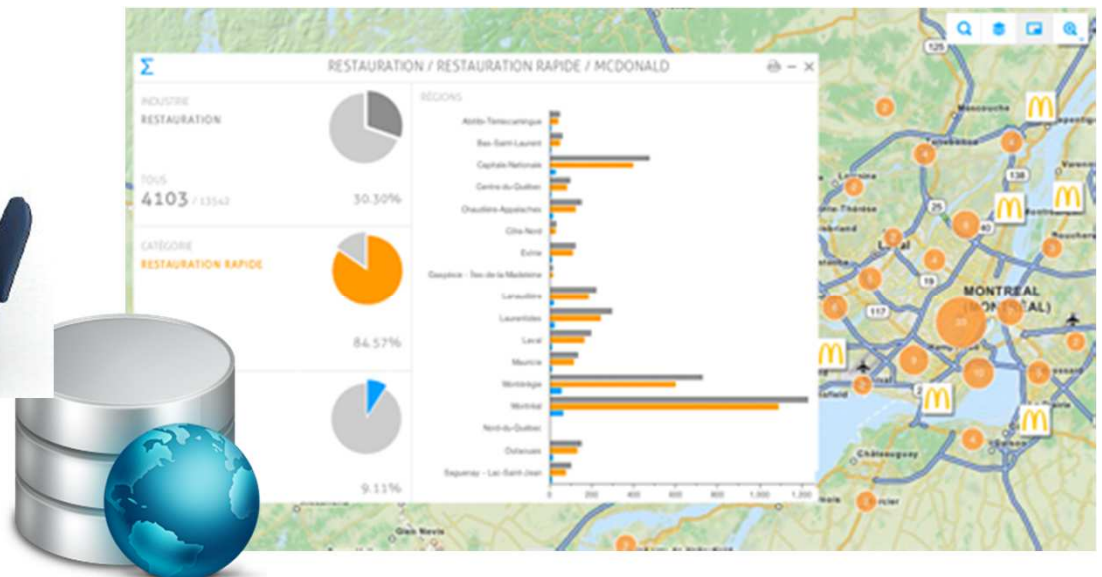
GeoBusinnes Intelligence (1/3)

Les technologies GeoBI permettent l'analyse des gros volumes de données avec une dimension spatiale :

Gestion → Analyse → Visualisation

Différents domaines d'application

Marketing, Santé, Environnement, ... et Agriculture





GeoBusinnes Intelligence (2/3)

Objectif

Améliorer les performances décisionnelles de l'entreprise

Comment ?

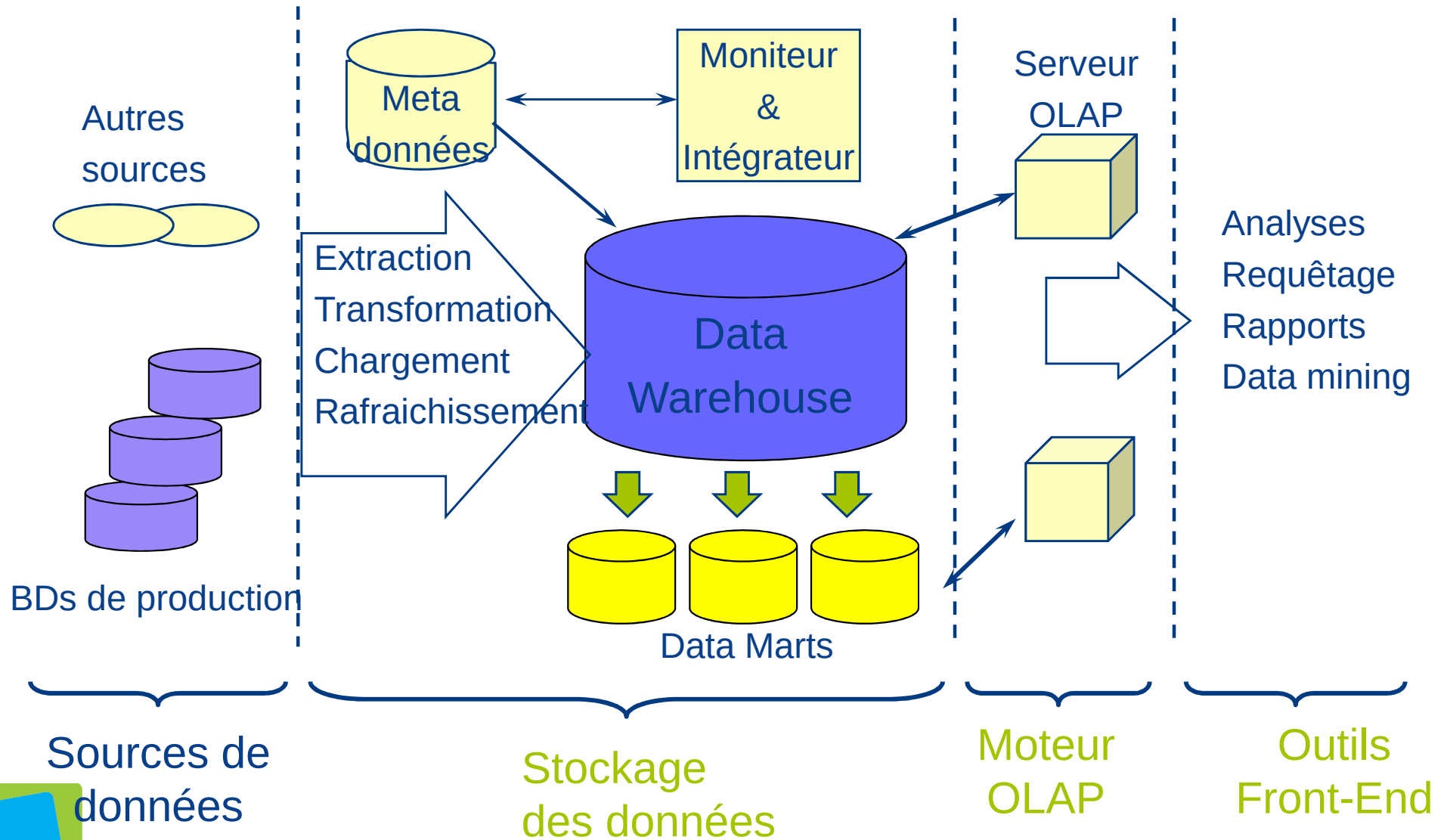
En répondant aux demandes d'analyse des décideurs

Exemple

Clientèle : « Qui sont mes clients ? »

« Ces clients sont-ils intéressants pour moi ? »

GeoBusinnes Intelligence (3/3)





GeoBusinnes Intelligence : Data Warehouse et OLAP (1/2)

«Un Data Warehouse (Entrepôt de données) est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision.»

Caractéristiques :

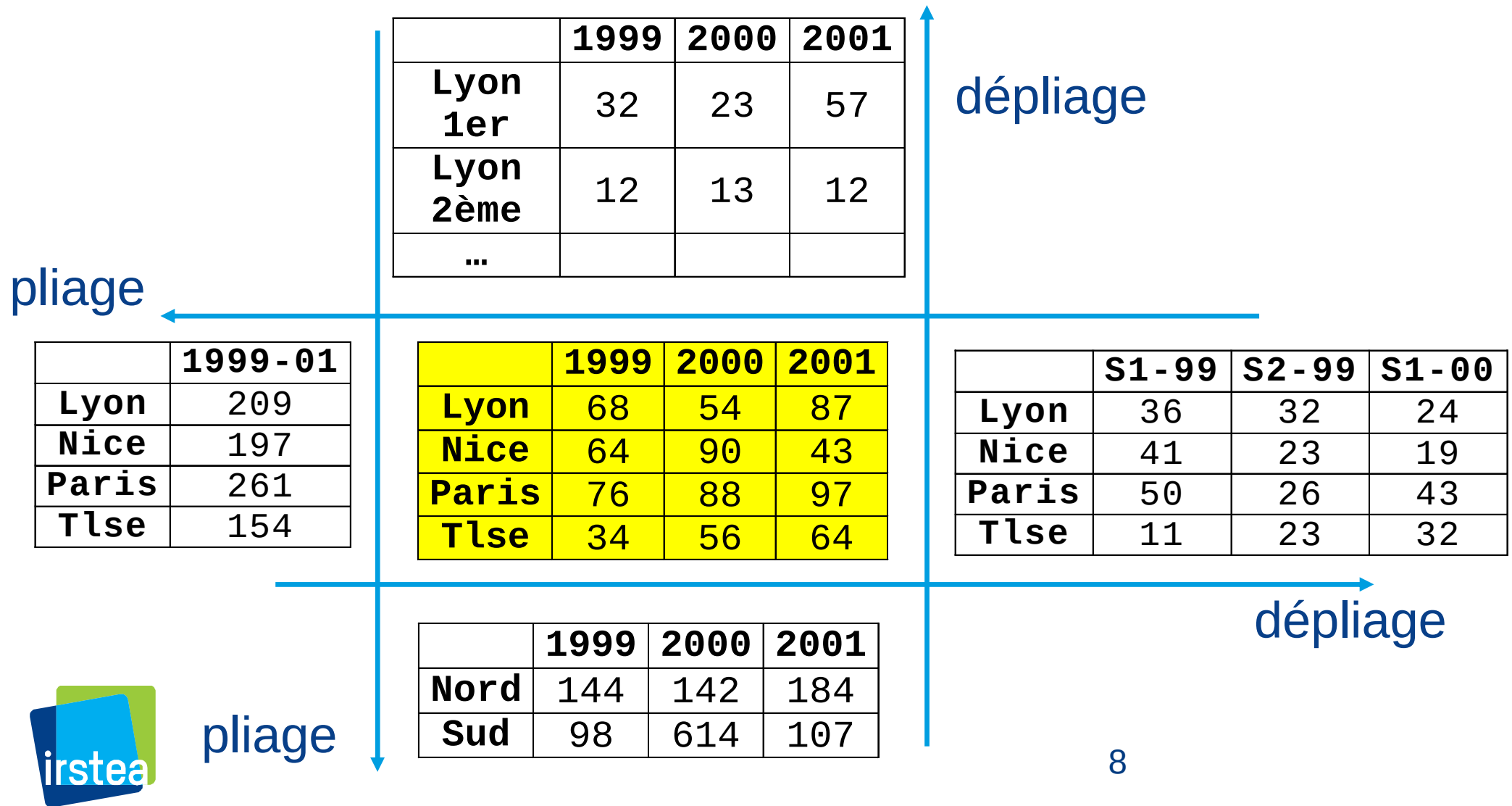
Orientation sujets («métiers»)

Données intégrées

Données non volatiles

Données datées

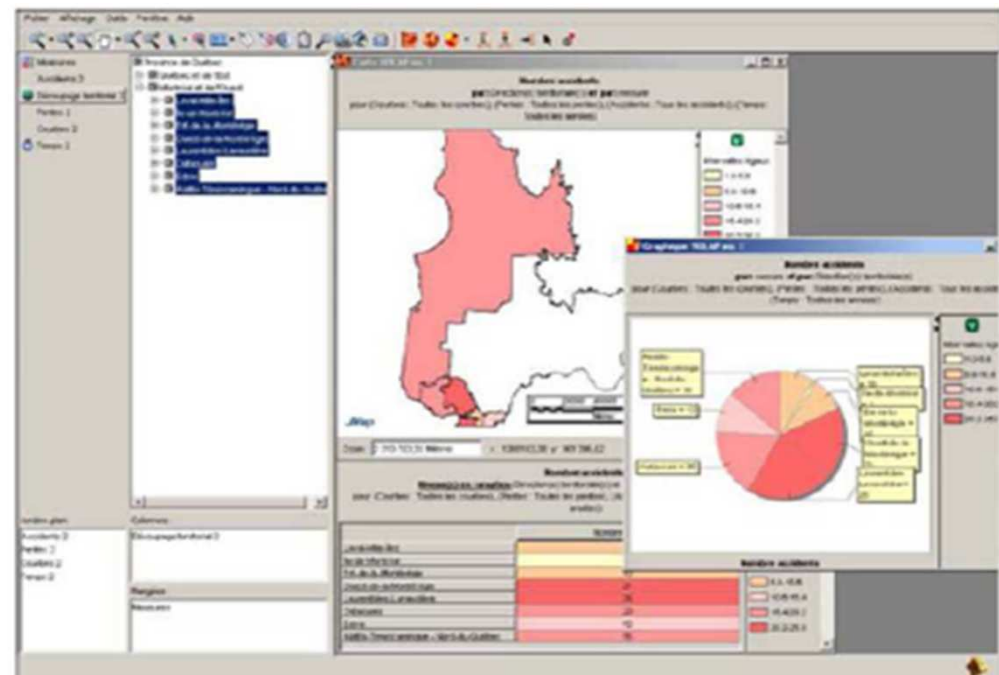
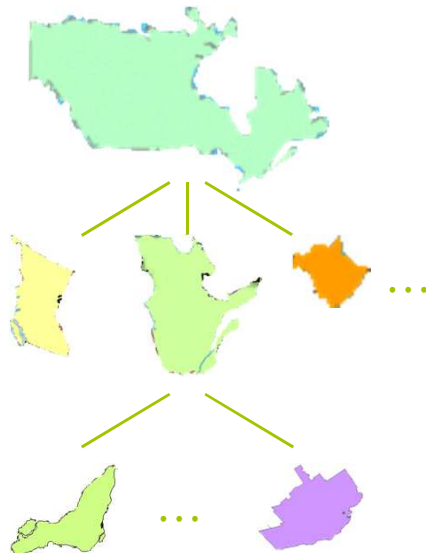
GeoBusinnes Intelligence : Data Warehouse et OLAP (2/2)



GeoBusinnes Intelligence : OLAP Spatial (SOLAP)

« Une plate-forme visuelle spécialement conçue pour supporter l'analyse et l'exploration spatio-temporelles rapides et faciles des données multidimensionnelles composées de plusieurs niveaux d'agrégation à l'aide d'affichages cartographiques aussi bien qu'à l'aide de tableaux et diagrammes statistiques »

Dimension spatiale
géométrique



GeoBusinnes Intelligence : Reporting (1/2)

Définition d'indicateurs décisionnels à partir des requêtes SQL sur systèmes de bases de données classiques (OLTP)

Affichage avec des tableaux de bord statiques

Ex: Quelle est la consommation de fuel pour chaque jour?

Base de données
historique

-BDH-



Indicateur 1

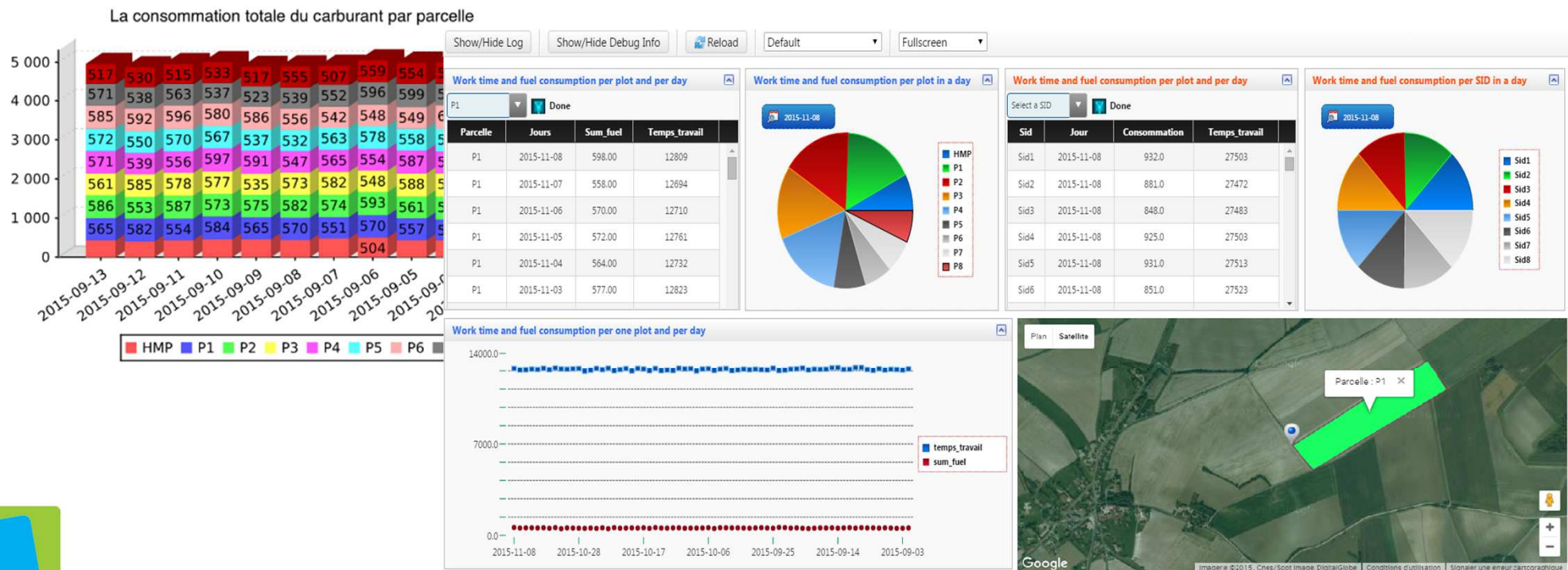
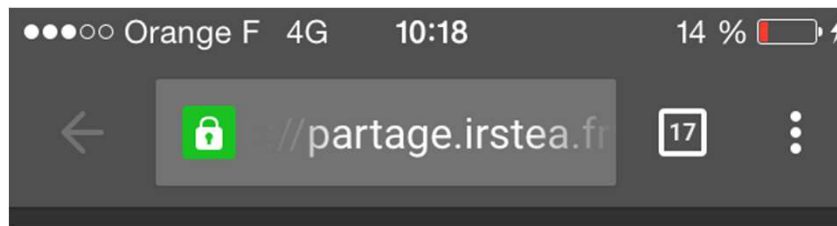
Indicateur 2

Indicateur 3



GeoBusinnes Intelligence : Reporting (2/2)

Affichage via sites web classiques ou mobiles





GeoBusinnes Intelligence : Stream

Un stream est une collection de données en temps réels et volatiles

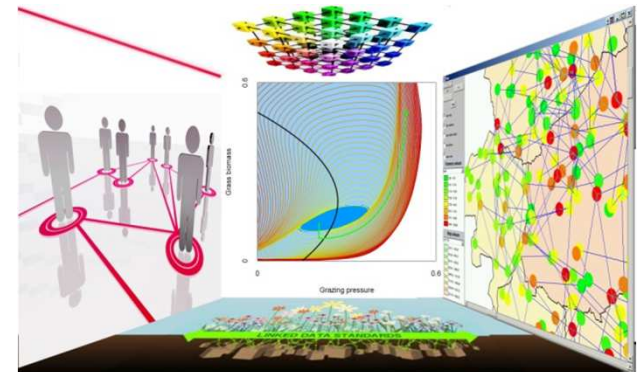
Un stream n'est pas stocké

Les requêtes sont continues

Adapté aux systèmes d'alertes

Ex: Alerte moi quand la température dépasse 45°

Intégration des données capteur dans les systèmes GeoBI





Les données des capteurs

Les capteurs ne présentent pas un système de gestion de données avancé car ils sont limités en ressources

Stockage de base (e.g. Fichiers txt, etc.)

Peu d'analyses possibles (e. g. données pas filtrées et corrigées)

Données très détaillées (e.g. données pas agrégées)

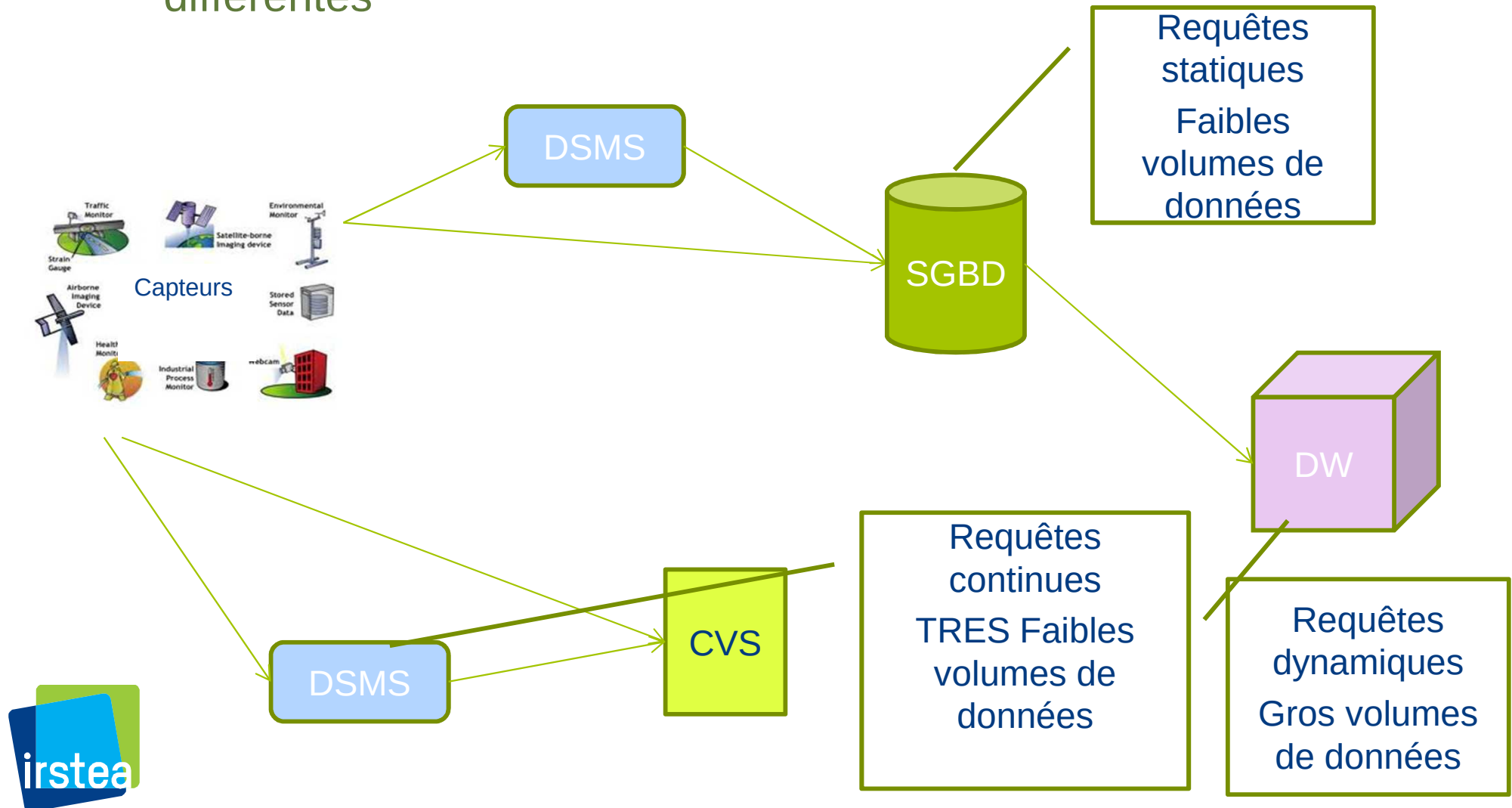
La communication des données capteur en milieu agricole peut être difficile

Problèmes de qualité (données incomplètes, etc.)

Données spatio-temporelles

Intégration des données capteur dans le GeoBI (1/4)

L'analyse des données capteur peut être faite avec des approches différentes



Intégration des données capteur dans le GeoBI (2/4)

Analyse

Indicateurs classiques aux échelles *exploitation* et *campagne*



Indicateurs fins au niveau de la position des capteurs (e.g. parcelle)
(Échelle spatiale)

Indicateurs fins au niveau de la fréquence d'acquisition (e.g. minute)
(Échelle temporelle)

Analyse spatiale (et multidimensionnelle) des indicateurs (e.g. *consommation moyenne par tracteur dans la parcelle P1*)





Intégration des données capteur dans le GeoBI (3/4)

Univers d'analyse

Déploiement à la demande

Automatisation de l'acquisition et donc alimentation automatique

Nouvelles sources de données (car les données peuvent être collectées aussi où l'utilisateur n'a pas toujours accès)



Intégration des données capteur dans le GeoBI (4/4)

Qualité

ETL et Data Warehouse permettent la résolution des problèmes de qualité (e.g. interpolation spatiale) et de la masse de données (e.g. indexes, vues matérialisées)



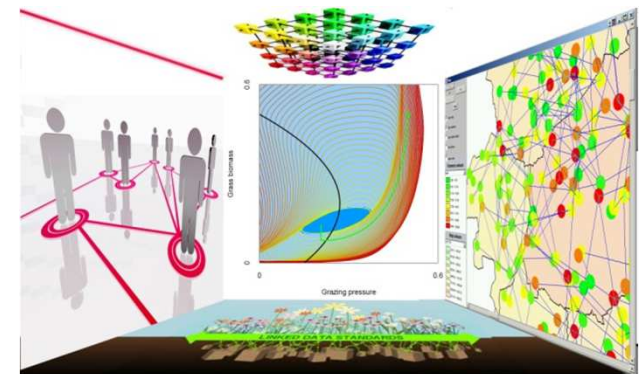
ETL Spatial

ETL : Prise en compte de l'information spatiale (e.g. inclusion spatiale – Spatial Data Integrator) pour la définition des données multidimensionnelles

Pas seulement pour l'affichage cartographique des dimensions spatiales

Exemple

- Thème de Recherche MOTIVE





Le projet EDEN

Déploiement de capteurs pour l'acquisition de données de carburant

Définition d'un entrepôt de données basé sur les indicateurs de pilotage et ACV

Mise en place du processus de ETL

Le projet EDEN : données capteurs

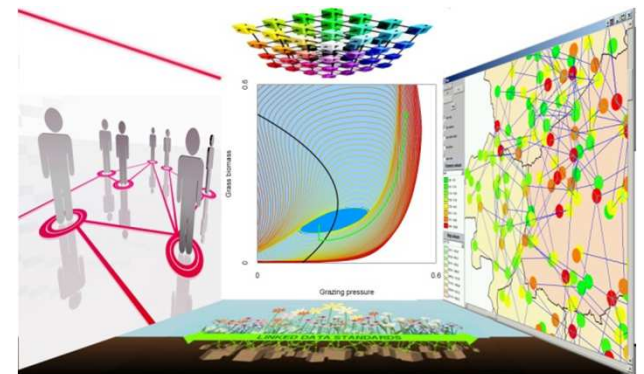


Représentation graphique des points enregistrés lors du transport et du travail de déchaumage

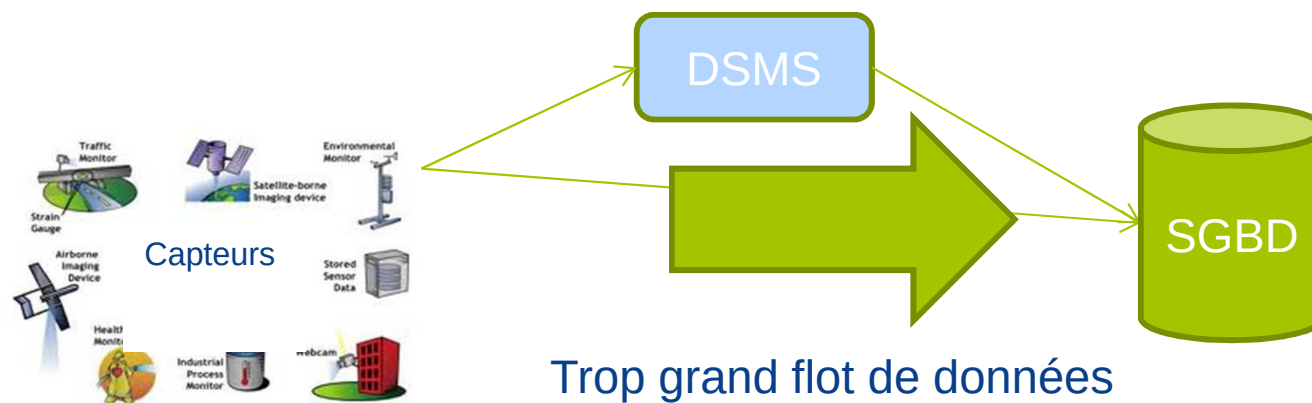
Importation des données du fichier "63450N_07_08_2013.LOG" dans le logiciel SIG Agromap (Agrocom CLAAS)

Points enregistrés lors d'actions sur le piton d'attelage et sur le relevage du tracteur

Et le Big Data?



GeoBI et Big Data





Problèmes : Avant l'analyse (1/2)

Performance :

Grands flots de données ➡ Les SGBD relationnels (SQL) n'arrivent pas à assurer l'insertion de données (mécanismes internes de contrôle des données)

Solution du Big Data : les systemes NoSQL

MongoDB et Cassandra permettent l'insertion de données (write) de multiples sources

Ex: MongoDB 35000 insert per second depuis 14 clients (*van der Veen et al. 2012*)

Problèmes : Avant l'analyse (2/2)

Structure de données :

Données non structurées (selon le modèle SQL) issues de capteurs

message_id bigint	numero_serie character varying(50)	message character varying(256)	date timestamp without time zone
1000008	vbox-001-00002	GPS:1443795701;162141021015;4.5442079E;48.5434526N	2015-10-02 16:21:41.0599
1000009	vbox-001-00003	GPS:1443795701;162141021015;4.5452079E;48.5404526N	2015-10-02 16:21:41.1531
1000010	vbox-001-00004	GPS:1443795702;162142021015;4.5462079E;48.5374526N	2015-10-02 16:21:42.4931
1000011	vbox-001-00004	MEMS:1443795760;0141D4935490813F	2015-10-02 16:21:42.4953
1000012	vbox-001-00004	MEMS:1443795746;0141D4935390013D	2015-10-02 16:21:42.4984



Problèmes : Avant l'analyse

Structure de données:

Données non structurées (selon le modèle SQL) issues de capteurs

Solution du Big Data : les systèmes NoSQL

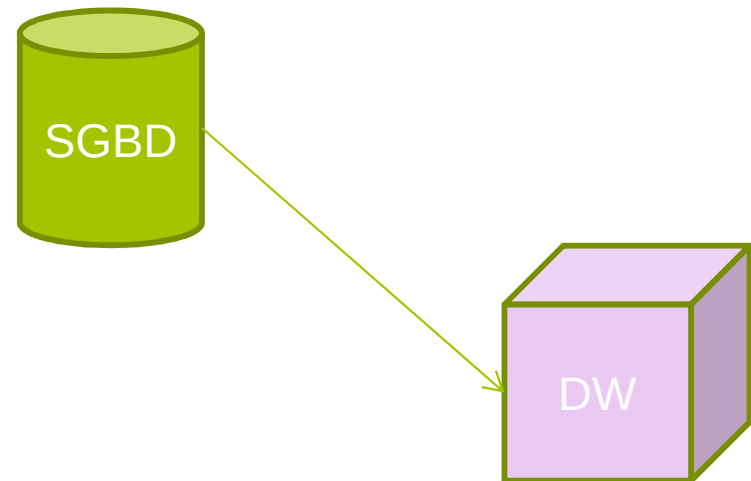
Schemaless: absence d'une structure de données prédéfinie

- Ex: MongoDB

```
db.sensDW.insert({  
  sens:"TracteurJD",  
  Time:"19h00-12-01-2016"  
  Info:"GPS45,6-7,0"});
```

```
db.sensDW.insert({  
  sens:"TracteurJD",  
  Time:"19h01-12-01-2016"  
  Info:"fuel10"});
```

GeoBI et Big Data





Problèmes : L'analyse

Gros volumes de données et vélocité

Solutions du Big Data :

Calcul parallèle (e.g. map-reduce)

Stockage reparté (e.g. hadoop)

Infrastructures publiques et privées (e.g. Amazon)



Problèmes : Quelles analyses ?

John Deere : « *Nous avons pris de l'avance par rapport à nos concurrents car nous avons développé notre propre système de GPS. Pour l'instant nous ne savons pas quoi en faire. Notre métier, ce n'est pas l'analyse de données, mais la construction de machines* »



Problème : Quelles analyses ?

Scenario existant : « Solutions vers Utilisateurs »

Les systèmes existants définissent un ensemble prédéfinis d'indicateurs

Un nouveau indicateur → nouveau système de GeoBI, capteurs, etc.

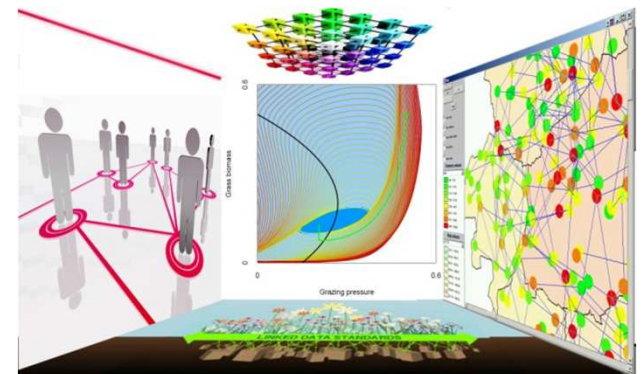
Scenario envisagé: « Utilisateurs vers Solutions »

1. Quels sont les besoins de l'utilisateur (type de requêtes) ?
(Reporting, SOLAP, etc.)
2. Big data ou Small Data?

> Conception du système d'acquisition

> Conception du system de GeoBI

Conclusions





Conclusions

Les technologies de GeoBI permettent des analyses différentes des données issues de capteurs

Les nouvelles technologies du Big Data représentent des solutions effectives pour l'analyse (GeoBI) de grandes quantités de données issues de capteurs

Conception des application de GeoBI pour Big Data en fonction des besoins spécifiques des utilisateurs finaux : du capteur au tableau de bord



GeoBusinnes Intelligence : ETL (1/2)

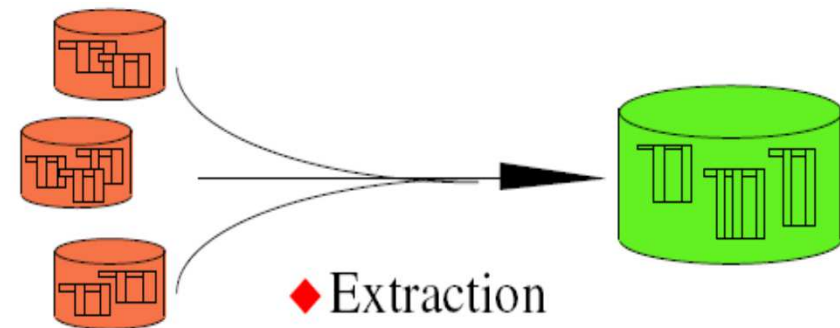
Extraction, Transformation et Chargement des données

Besoin d'outils spécifiques pour :

accéder aux bases de production (requêtes sur des BD hétérogènes)

améliorer la qualité des données : «nettoyer», filtrer, ...

transformer les données : intégrer, homogénéiser, ...



◆ Extraction

◆ Transformation

- filtrer
- trier
- homogénéiser
- nettoyer
- ...

◆ Chargement
(Loading)

GeoBusinnes Intelligence : Data Warehouse et OLAP (3/6)

Dimensions

Représentent les axes d'analyses (produit, localisation, temps...)

Description textuelle des membres

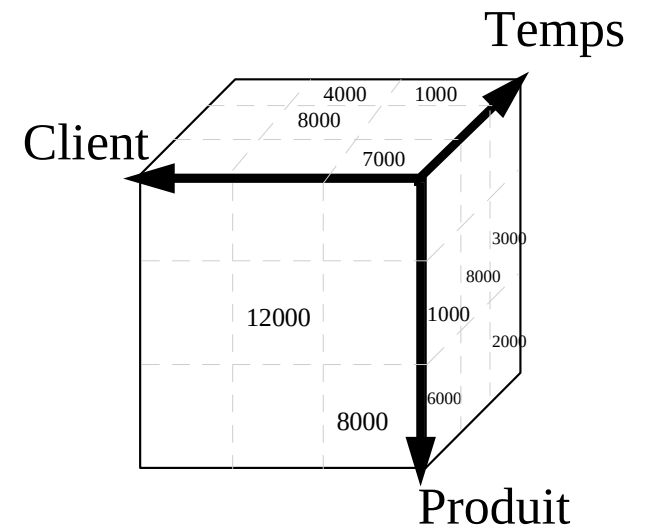
Organisées en hiérarchie (ville < département < région < pays)

Faits

Sujet de l'analyse

Décrits par des valeurs numériques (mesures)
(ventes, température, etc.)

Mesures sont agrégées avec des fonctions
mathématiques (somme, min) sur les hiérarchies

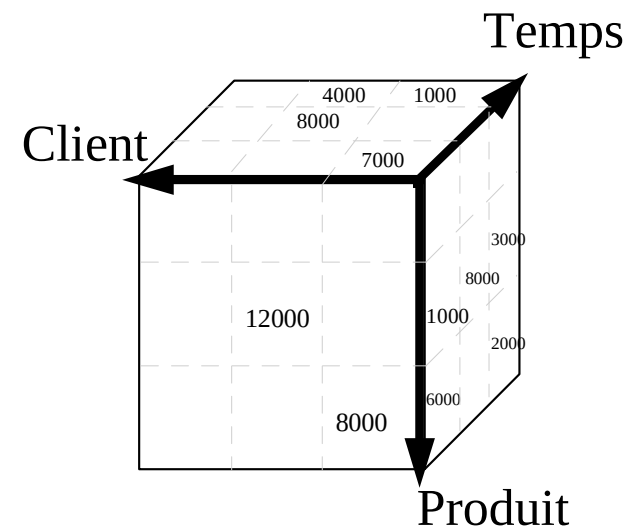


GeoBusinnes Intelligence : Data Warehouse et OLAP (3/6)

Un Indicateur est donc une mesure et une fonction d'agrégation

Un Indicateur est analysé selon différents axes d'analyse et à différentes échelles

avec les opérateurs OLAP





GeoBusinnes Intelligence : OLAP vs Data Mining

Data Mining

L'utilisateur cherche des corrélations non évidentes

Peu de données, une table

Prédiction

OLAP

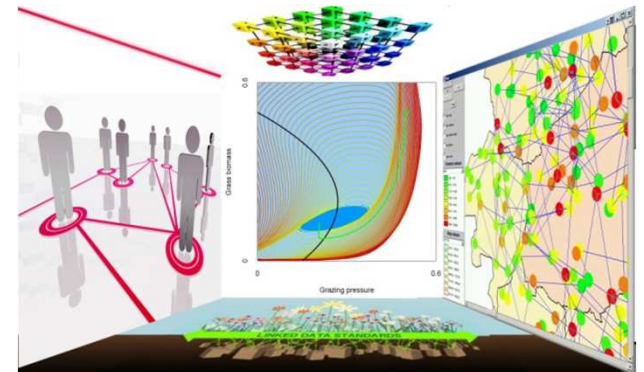
L'utilisateur cherche à confirmer des intuitions

Gros volumes de données et multi-tables

Exploration et agrégation

OLAP et après Mining...

Le projet DISP'EAU

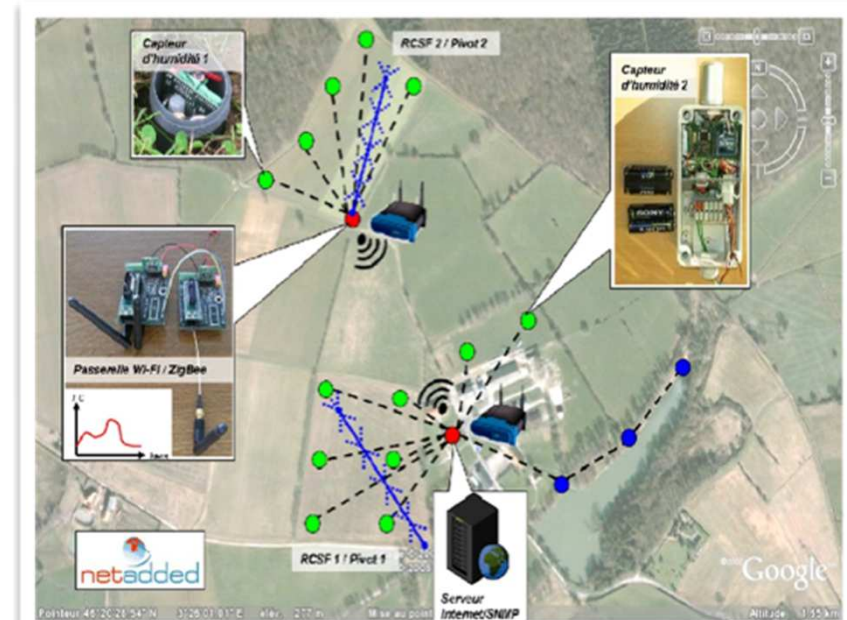
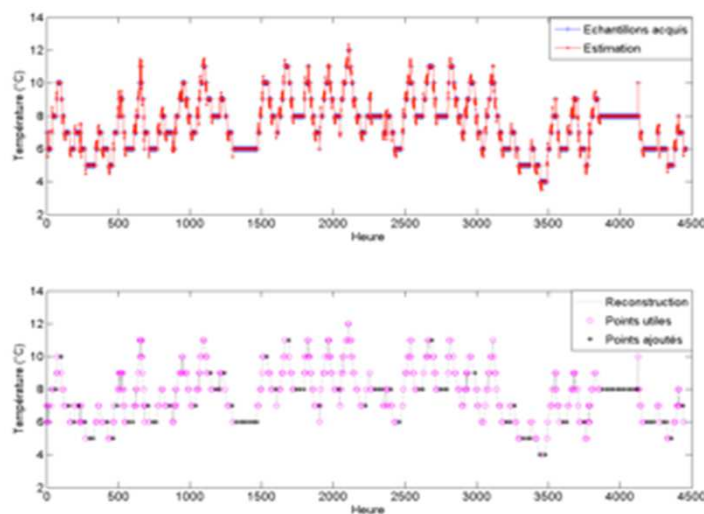


Le projet DISP'EAU (1/2)

Mise en œuvre de modèle de données pour limiter l'envoi d'information

Projet Européen NetAdded pour le suivi de l'irrigation

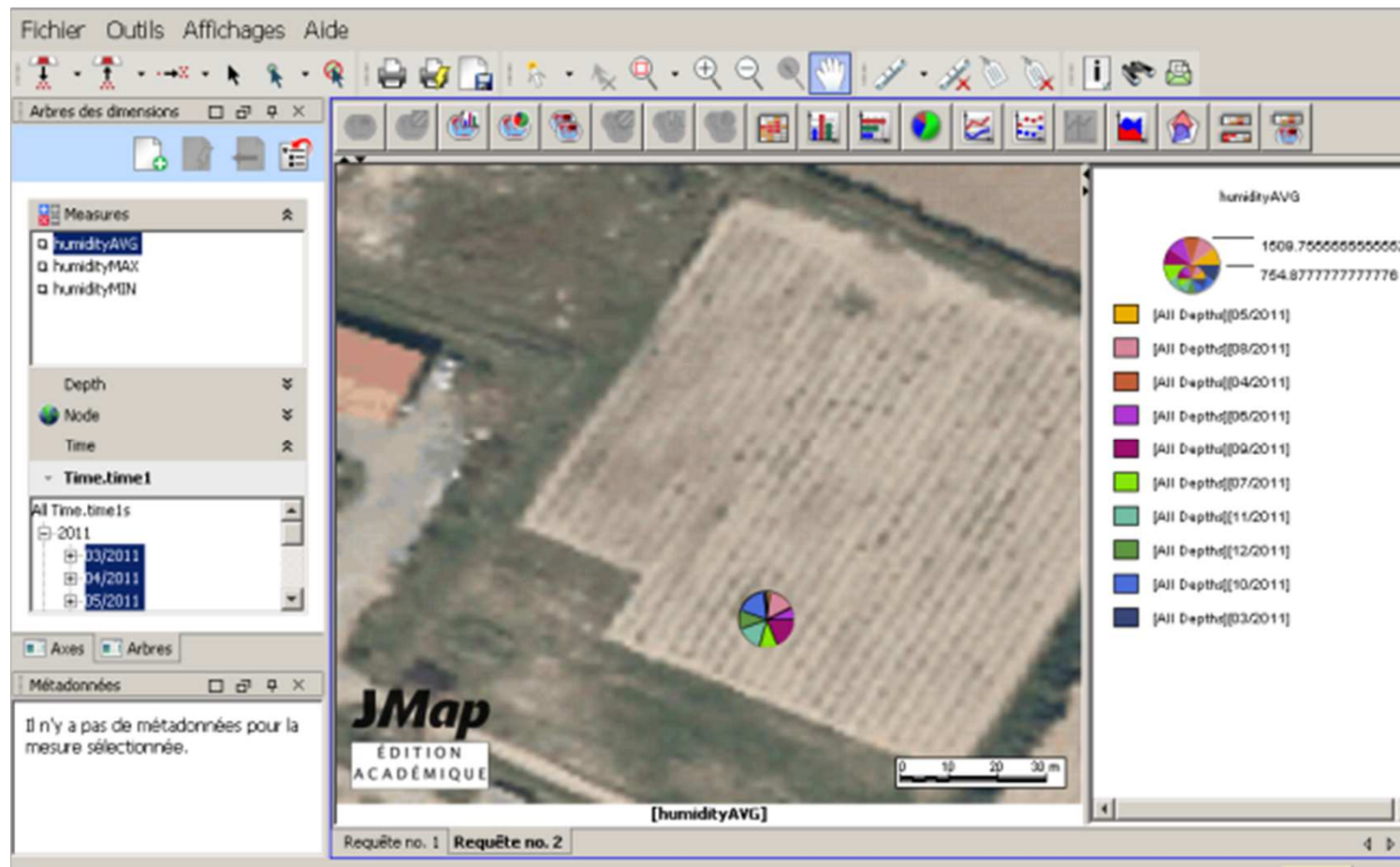
Suivi de la température, de l'humidité du sol



Le projet DISP'EAU (2/2)

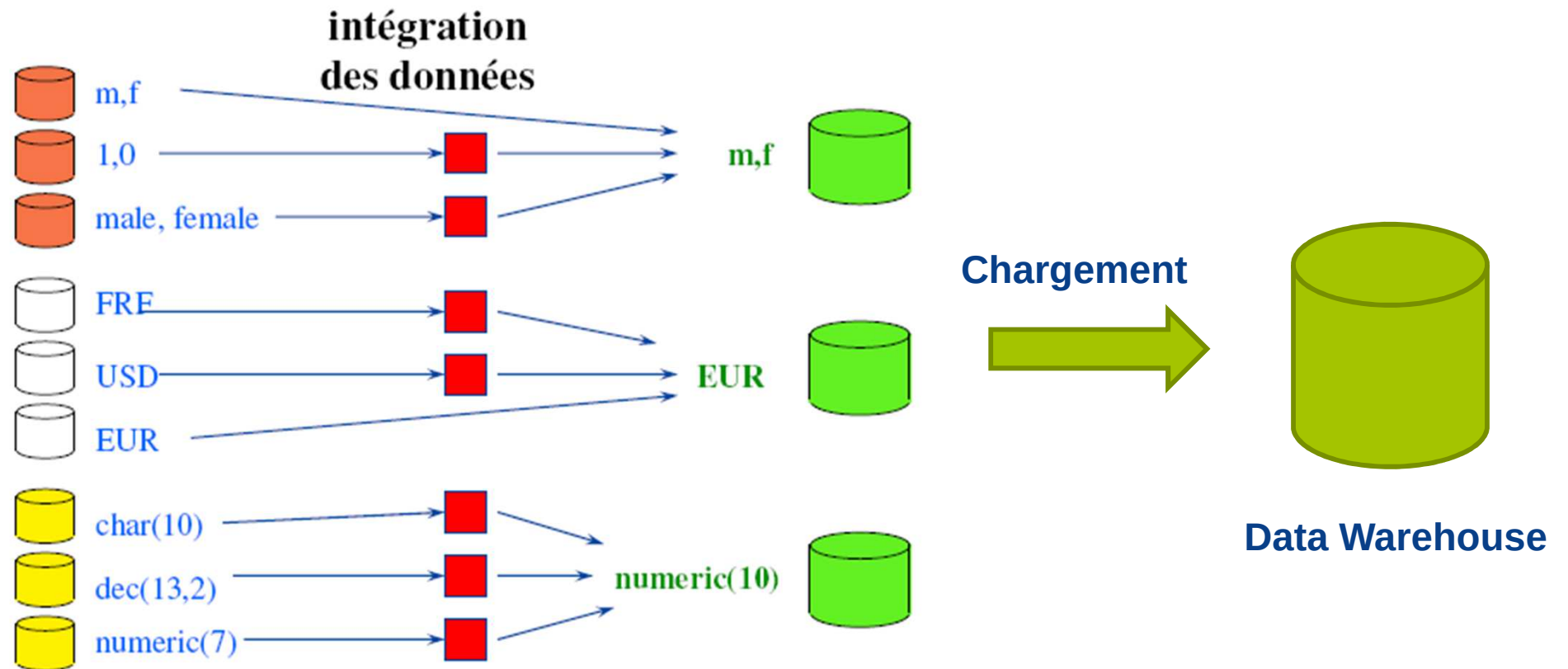
Indicateurs avec l'outil SOLAP

Un exemple : humidité moyenne par capteur et mois ?

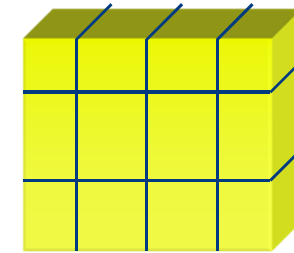
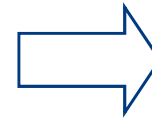
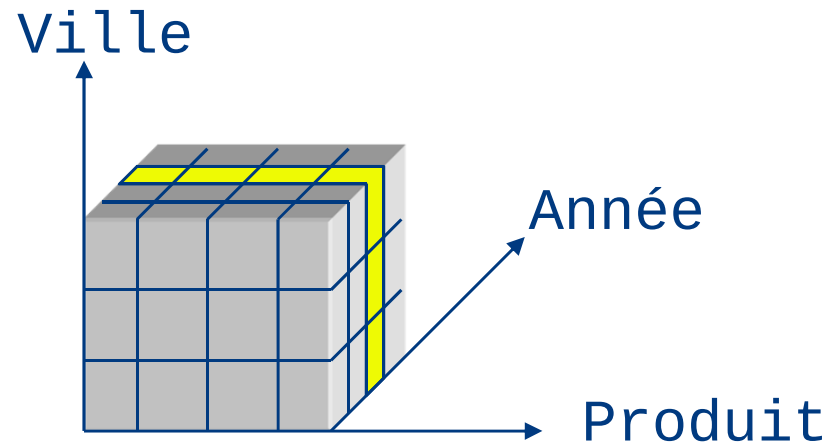


GeoBusinnes Intelligence : ETL

Exemple de transformation



GeoBusinnes Intelligence : Data Warehouse et OLAP (4/6)



		1999	2000	2001
Frais	Lyon	678	354	54
	Nice	765	245	76
	Paris	788	223	555
	Tlse	543	763	556
Liquide	Lyon	543	876	676
	Nice	...		
...	...			

		1999
Frais	Lyon	678
	Nice	765
	Paris	788
	Tlse	543
Liquide	Lyon	543
	Nice	...
...	...	

GeoBusinnes Intelligence : Data Warehouse et OLAP (5/6)





Intégration des données capteur dans le GeoBI (2/5)

L'intégration des données de capteurs dans le systèmes GeoBI permet une analyse des indicateurs à différentes échelles spatiales et temporelles

Les systèmes GeoBI permettent de « résoudre » des problématiques liées aux données de capteurs

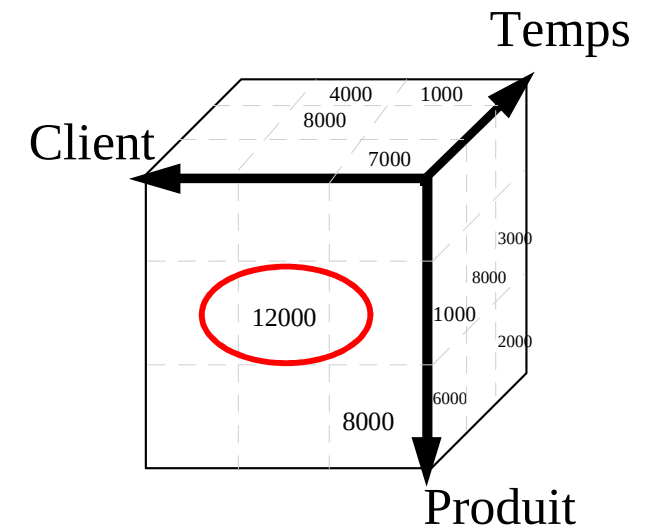
Le couplage de ces deux systèmes offre des nouvelles possibilités d'analyse des indicateurs

GeoBusinnes Intelligence : Data Warehouse et OLAP (2/3)

Évolution du montant total des ventes des produits de type « alimentaire » dans le magasin de Lyon sur les 3 dernières années ?

Un modèle multidimensionnel

- ⇒ Requête = naviguer dans le cube
- ⇒ **Visualiser/calculer un indicateur**



GeoBusiness Intelligence (3/4)

