

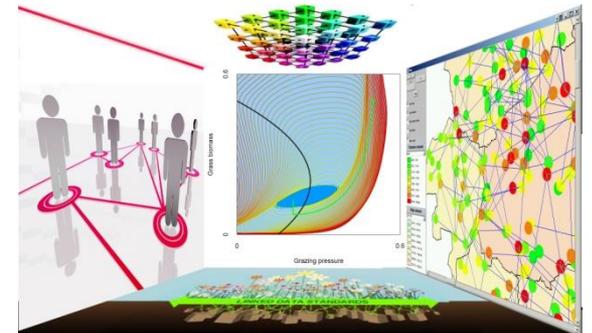


ProtOLAP : Suite pour la conception de cubes OLAP

Sandro Bimonte, *Irstea, France*



www.irstea.fr



Qui nous sommes

Equipe COPAIN, Irstea (bientôt INRA)

- 16 Permanents chercheurs et ingénieurs
- Unité de Recherche TSCF « Technologies et Systèmes d'information pour les agrosystèmes »
- Clermont Ferrand

Compétences :

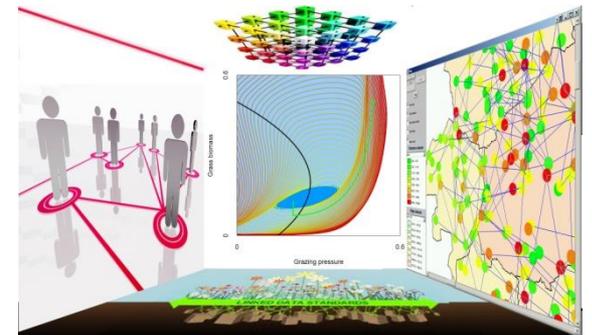
- Systèmes d'Information, (Geo)Business Intelligence, Systèmes d'Information Géographique
- Capteurs, Réseaux de capteurs sans fils
- Ontologies, Linked Open Data, Gestion des connaissance



Plan

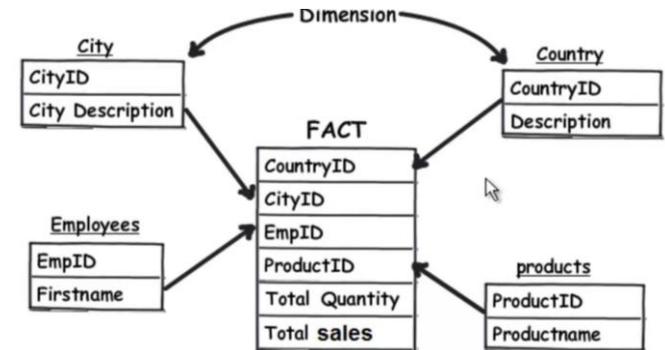
- Motivation
- ProtOLAP methodologie
- ProtOLAP suite
- Conclusion

Motivation



Préliminaires

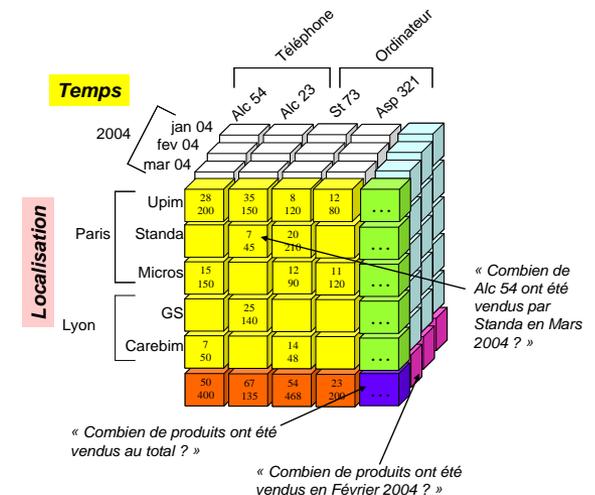
Un **entrepôt de données** est « un schéma multidimensionnel de base de données »



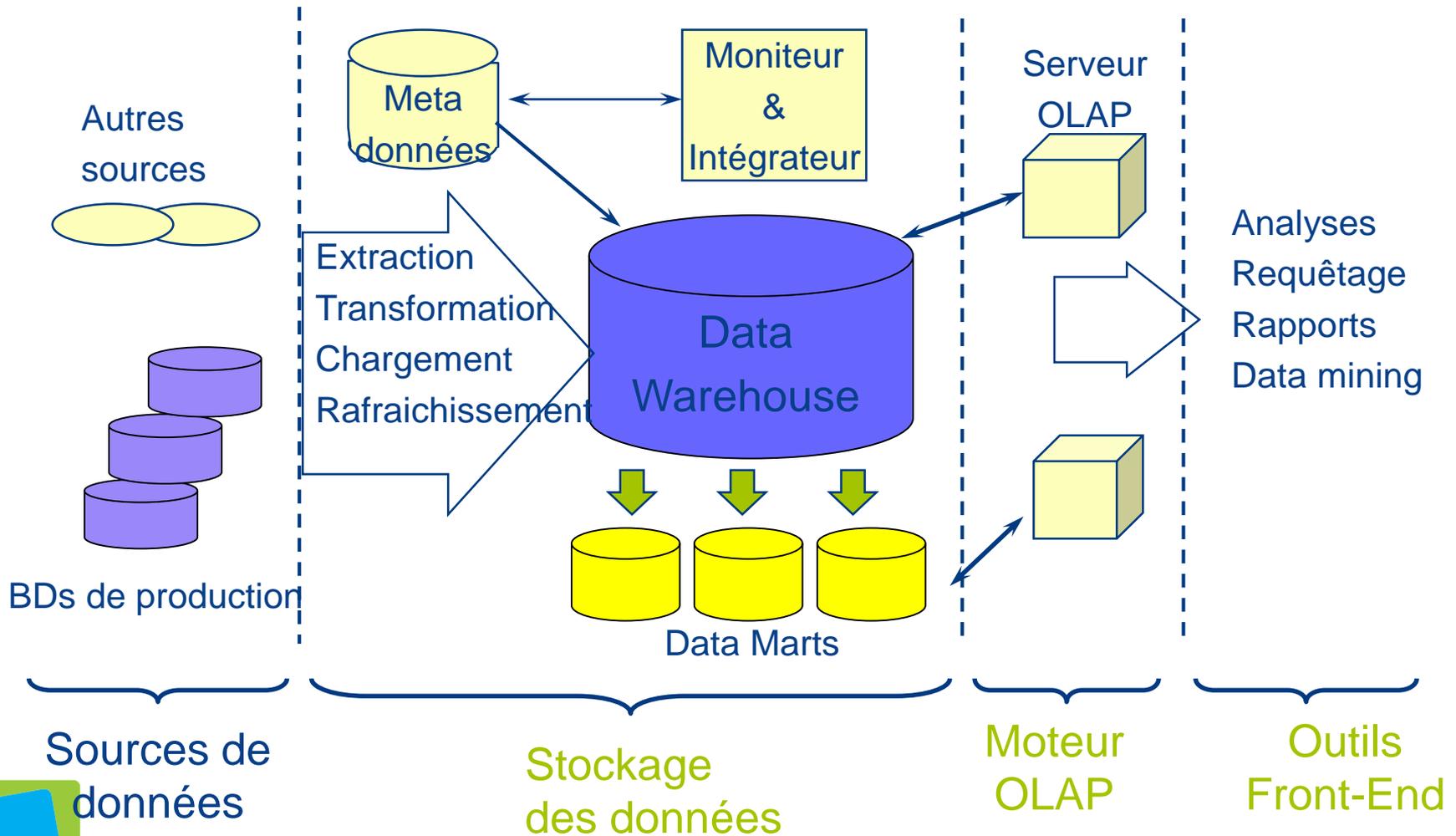
Un **indicateur** est « une mesure + une fonction d'agrégation »

Total sales + SUM

Un **Cube** est « l'ensemble des valeurs des indicateurs pour tous les niveaux de l'entrepot de données »



Préliminaires



Motivations : Décideurs



decision makers

La majorité des projets OLAP échoue car les cubes OLAP implémentés ne correspondent pas aux besoins de décideurs

Très souvent les décideurs n'ont pas de compétences en OLAP

Ils n'arrivent pas à exprimer et valider leurs besoins d'analyse en utilisant des formalismes complexes (UML, etc.)



Très souvent les indicateurs OLAP sont assez complexes
Ils n'arrivent pas à bien l'exprimer et les valider

Motivations : Experts OLAP



designers

Conception et maintenance des **systems d'information** avec les modèles conceptuels (ER, UML, etc.)



Très souvent pour la conception et la maintenance des systèmes de **Business Intelligence** les modèles conceptuels ne sont pas utilisés



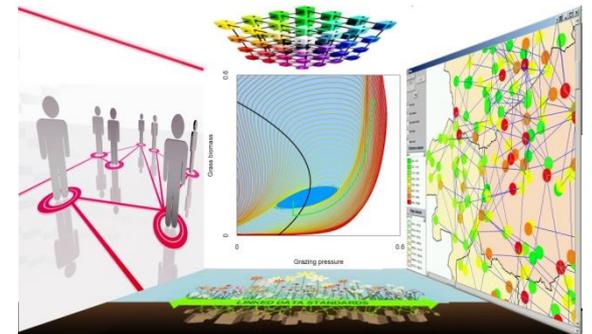
L'implémentation des indicateurs OLAP complexes est cause d'erreurs

Les formules implémentées ne correspondent pas aux attentes des décideurs

L'implémentation des formules complexes est difficile

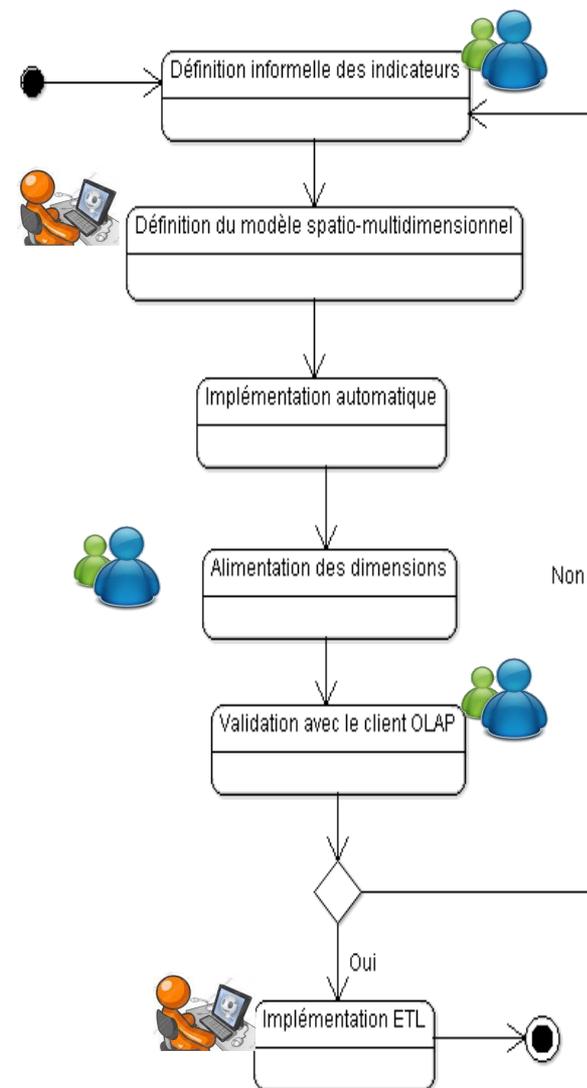


Méthodologie ProtOLAP



ProtOLAP : méthodologie (1/2)

1. Les décideurs expriment leurs besoins d'analyse en langage naturel
2. Les experts OLAP les traduisent dans un modèle UML
3. Le modèle UML est automatiquement implémenté dans un modèle logique pour SGBD et le serveur OLAP
4. Des données « sample » sont chargées dans le SGBD
5. Un prototype est livré aux décideurs
6. Les décideurs “jouent” avec le prototype



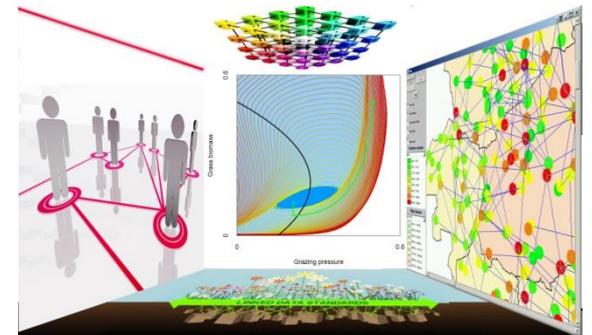
ProtOLAP : méthodologie (2/2)

1. Les décideurs “jouent” avec le prototype avec un client web OLAP
2. et avec un click une page wiki est créer/ouverte pour laisser un commentaire sur un problème rencontré (ex : nom dimension, niveau mal placé, etc.)

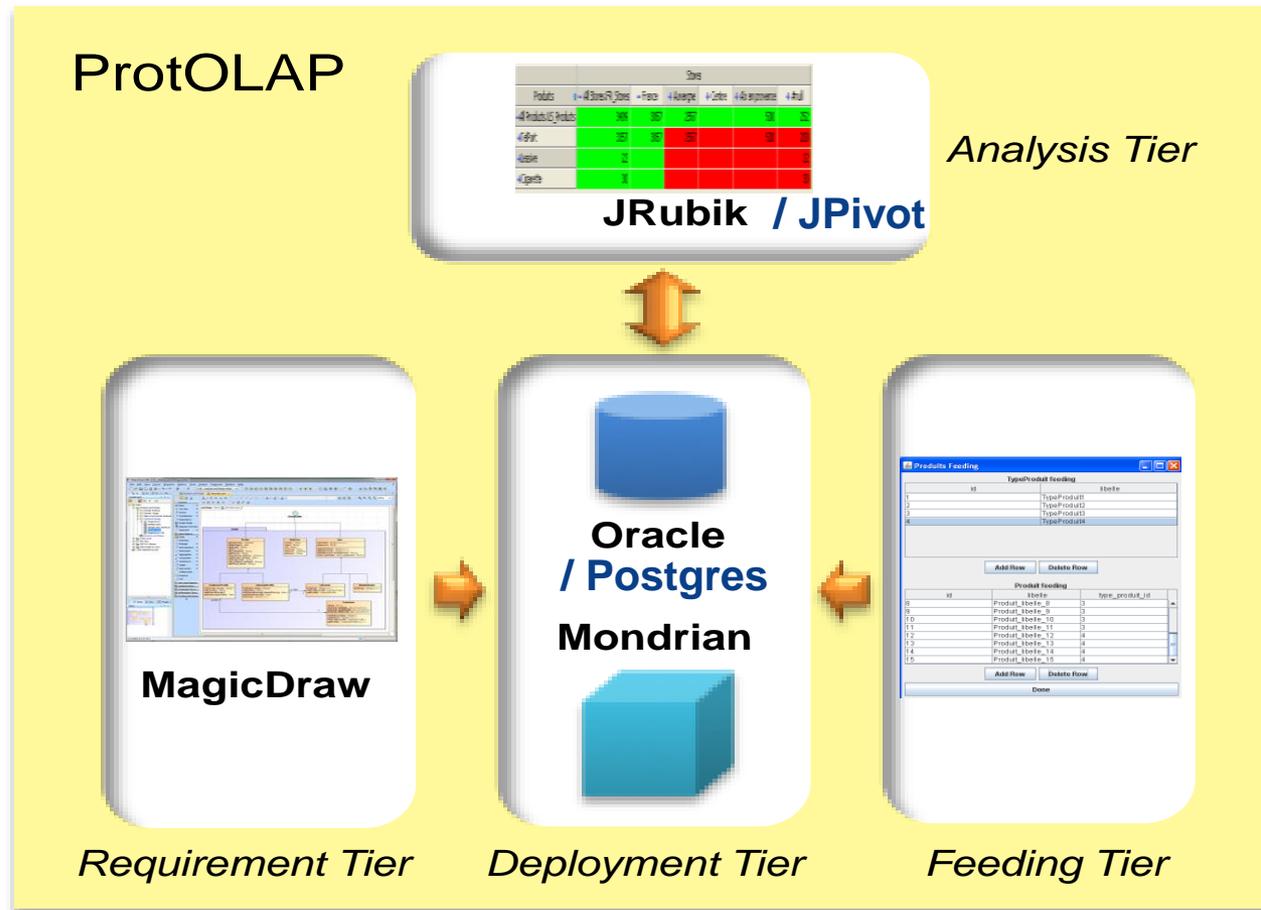
Nouveauté : une page wiki est associée à une requête OLAP (et pas à une table de base de données)

Permet aux décideurs et aux experts OLAP de communiquer à distance et de façon asynchrone sur tout élément du cube

ProtOLAP tool

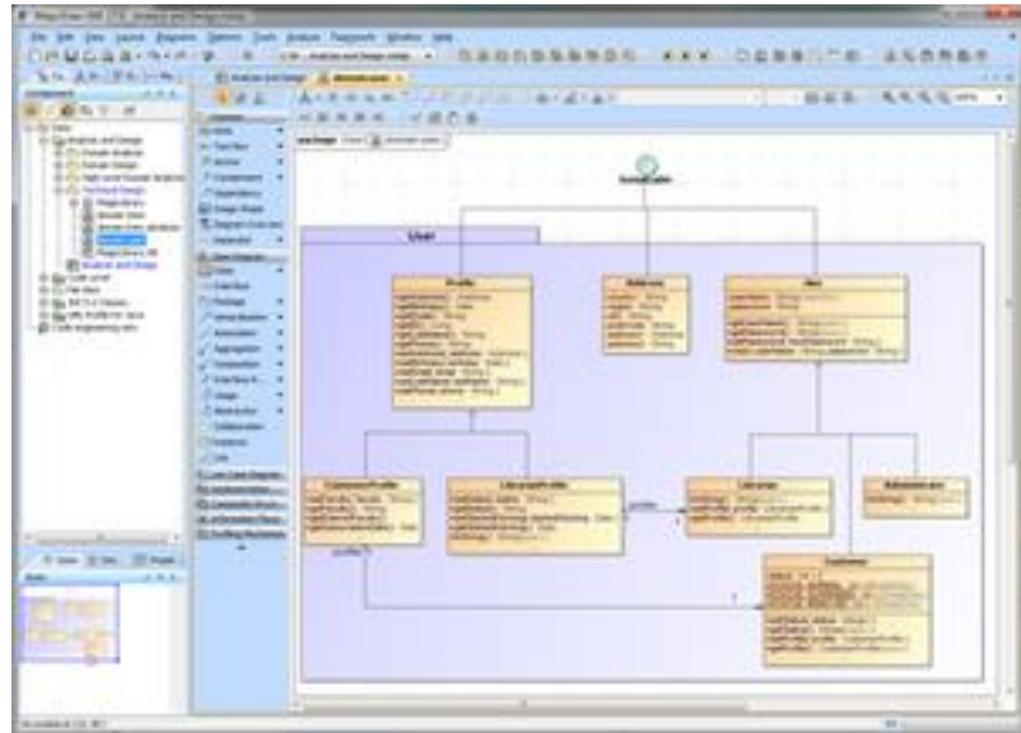


ProtOLAP : Architecture



ProtOLAP tool: Requirement tier

MagicDraw est un outil payant de modélisation UML (diagrammes de classes, activité, séquence, etc.) de NoMAGIC



ProtOLAP tool: Requirement tier

Comment représenter les cubes OLAP
(complexes) UML en MagidDraw?



Nous proposons ICSOLAP

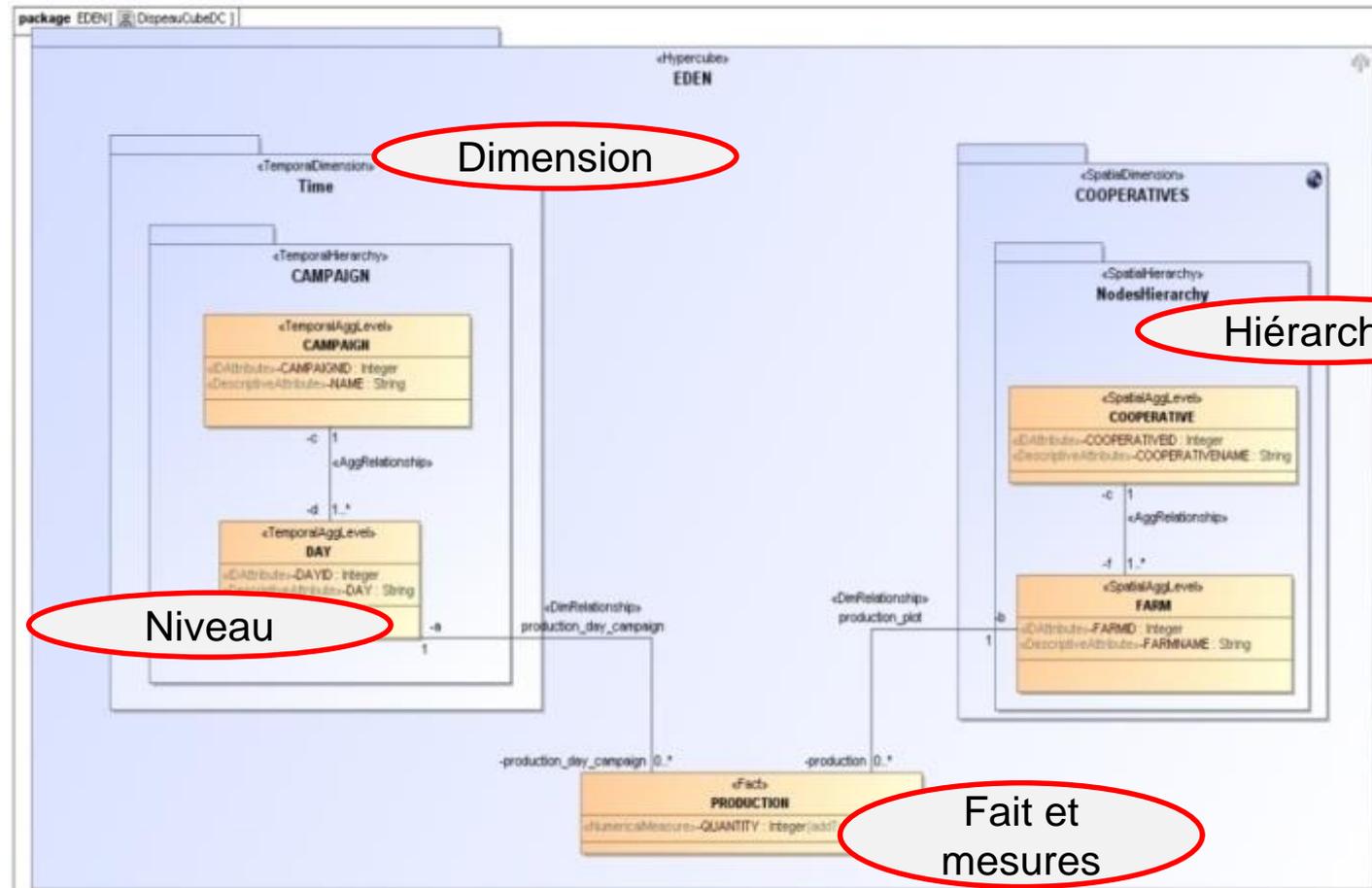


ICSOLAP est organisé en deux modèles

- Entrepot de données (faits-mesures et dimensions-hiérarchies)
- Agrégation (fonctions d'agrégation sur les mesures)

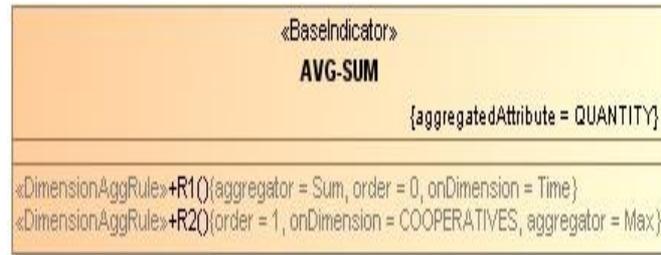
ProtOLAP tool: ICSOLAP exemple

Projet EDEN: Institut de l'élevage



ProtOLAP tool: ICSOLAP exemple

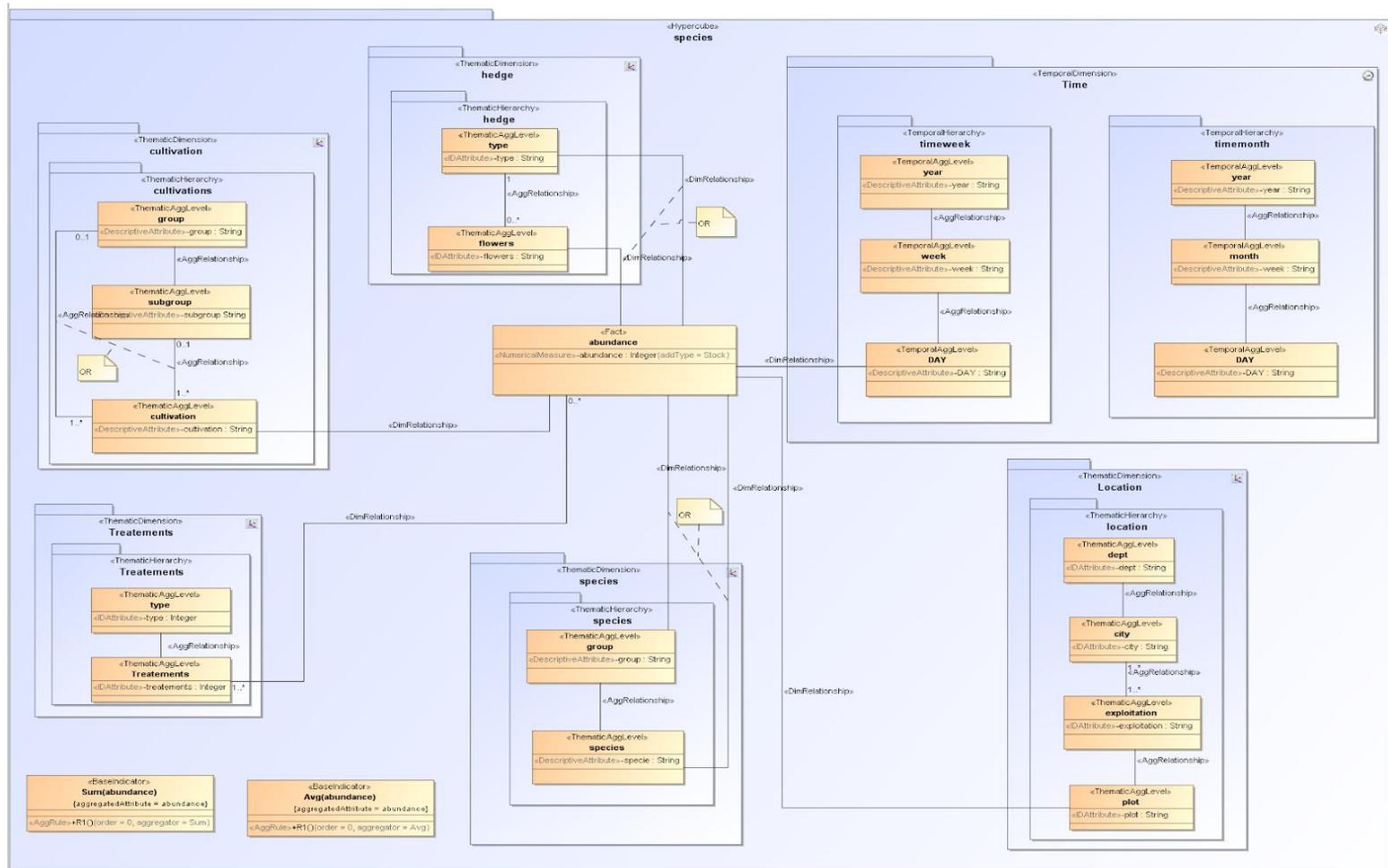
Projet EDEN: Institut de l'élevage



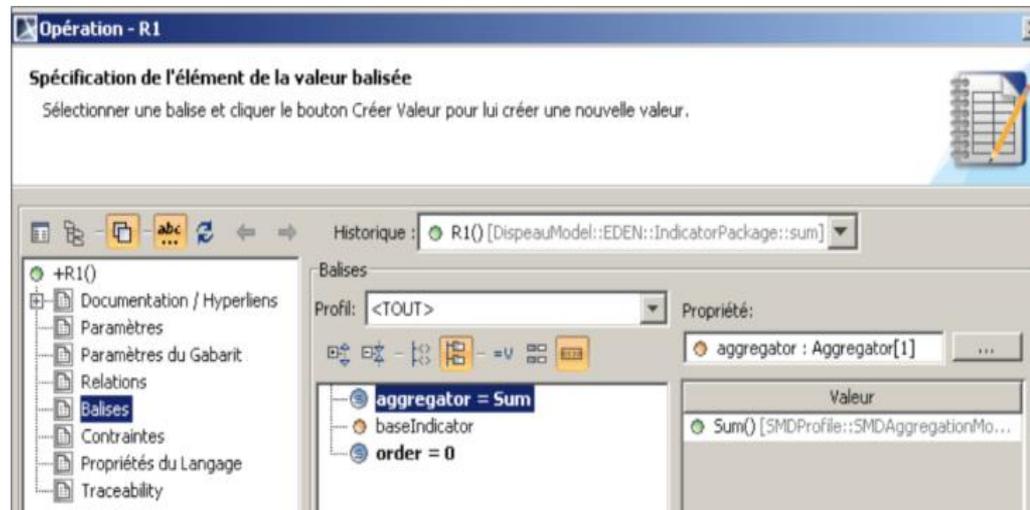
Total de lait produit par campagne et maximum par cooperative

ProtOLAP tool: ICSOLAP exemple

ICSOLAP permet de représenter toutes les structures complexes d'un entrepot de données (projet avec LPO et MNHN)



ProtOLAP tool: ICSOLAP exemple



ProtOLAP tool: Deployment Tier

- SGDBs supportés : ORACLE et Postgres

ProtOLAP utilise OCL2SQL (open-source)

ORACLE®

PostgreSQL

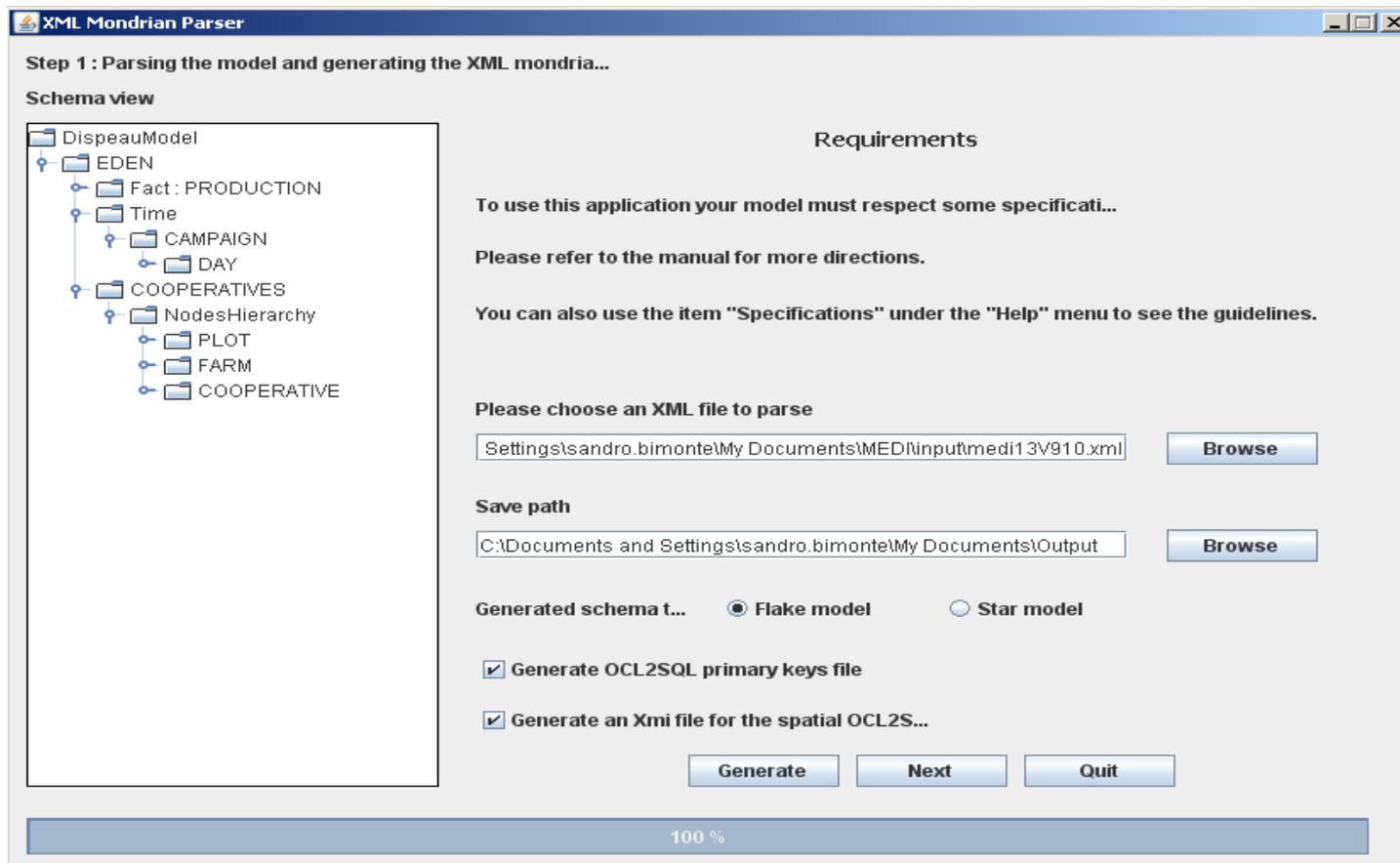


- Le serveur OLAP est Mondrian



ProtOLAP tool: Deployment Tier

1. Input modèle ICSOLAP UML du cube OLAP
2. Choix d'implémentation entre : "star schema" et "snowflake schema"

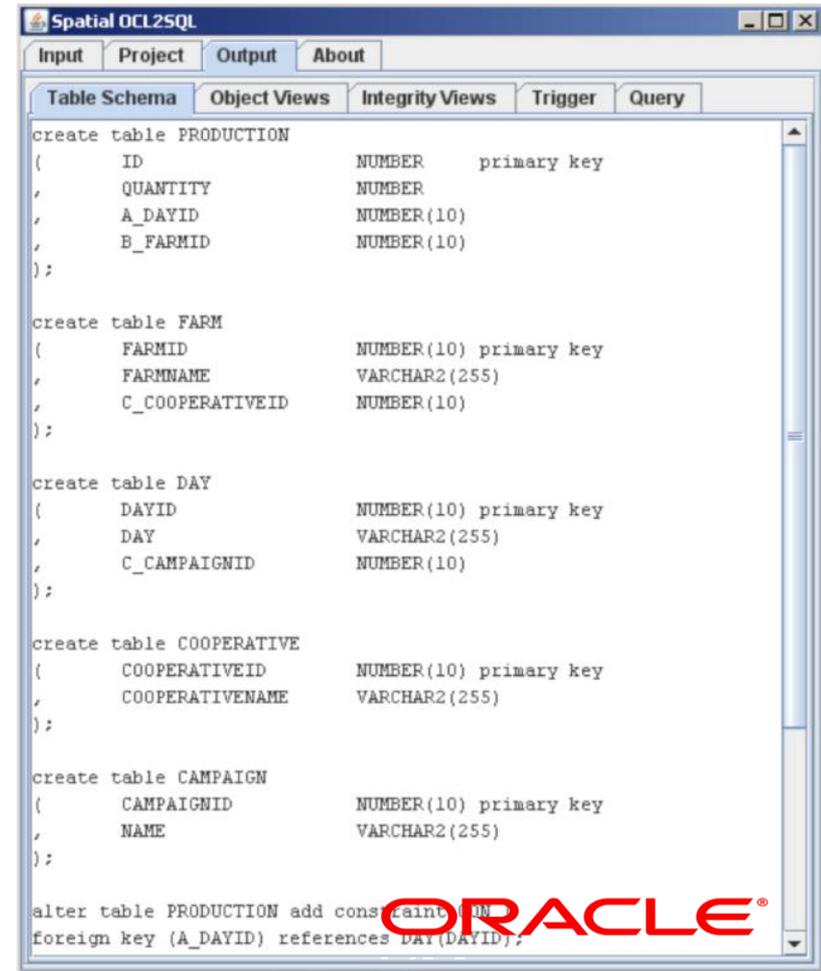


ProtOLAP tool: Deployment Tier

ProtOLAP génère

AUTOMATIQUEMENT

- Le schéma de l'entrepot de données dans le SGDB



The screenshot shows the Spatial OCI2SQL application window. The 'Table Schema' tab is active, displaying the following SQL code:

```
create table PRODUCTION
(
  ID                NUMBER          primary key
  QUANTITY          NUMBER
  A_DAYID           NUMBER(10)
  B_FARMID          NUMBER(10)
);

create table FARM
(
  FARMID            NUMBER(10) primary key
  FARMNAME          VARCHAR2(255)
  C_COOPERATIVEID  NUMBER(10)
);

create table DAY
(
  DAYID            NUMBER(10) primary key
  DAY              VARCHAR2(255)
  C_CAMPAIGNID     NUMBER(10)
);

create table COOPERATIVE
(
  COOPERATIVEID    NUMBER(10) primary key
  COOPERATIVENAME  VARCHAR2(255)
);

create table CAMPAIGN
(
  CAMPAIGNID       NUMBER(10) primary key
  NAME             VARCHAR2(255)
);

alter table PRODUCTION add constraint FK_PROD_DAY
foreign key (A_DAYID) references DAY(DAYID);
```

The Oracle logo is visible in the bottom right corner of the application window.

ProtOLAP tool: Deployment Tier

ProtOLAP génère **AUTOMATIQUEMENT**

- Le fichier XML Mondrian avec la définition du cube (mésures, dimension, etc.) avec le mapping sur le SGDB
- Le code MDX des indicateurs complexes (MDX est le langage standard des serveurs OLAP)

```
<!-- The Mondrian definition of the EDEN indicators -->
<Measure name="sum" column="QUANTITY" aggregator="sum" visible="false" formatString="Standard" />
- <CalculatedMember name="AVG-SUM" dimension="Measures" visible="true" formatString="#,###.##">
  <formula>Max(Descendants([COOPERATIVES].CurrentMember, [COOPERATIVES.NodesHierarchy].[FARM]),[Measures].
    [sum])</formula>
</CalculatedMember>
</Cube>
```

ProtOLAP tool: Feeding Tier

Une interface visuelle très simple permet d'insérer des données pour les niveaux des dimensions

- Manuellement
- Via un fichier CSV

COOPERATIVE feeding

COOPERATIVEID	COOPERATIVENAME
0	COOP_0

Add Row **Delete Row**

FARM feeding

FARMID	FARMNAME	c_COOPERATIVEID
0	Montoldre	0
1	Aubiere	0

Add Row **Delete Row**

Done

ProtOLAP tool: Feeding Tier

Une interface visuelle très simple permet de insérer les données des faits

Générateur automatique et manuelle des identifiants dimensions

Pour les mesures, il est possible définir une intervalle de valeurs

Fact table feeding step

QUANTITY	b_FARMID	a_DAYID
48	0	0
118	0	1
219	1	0
137	1	1

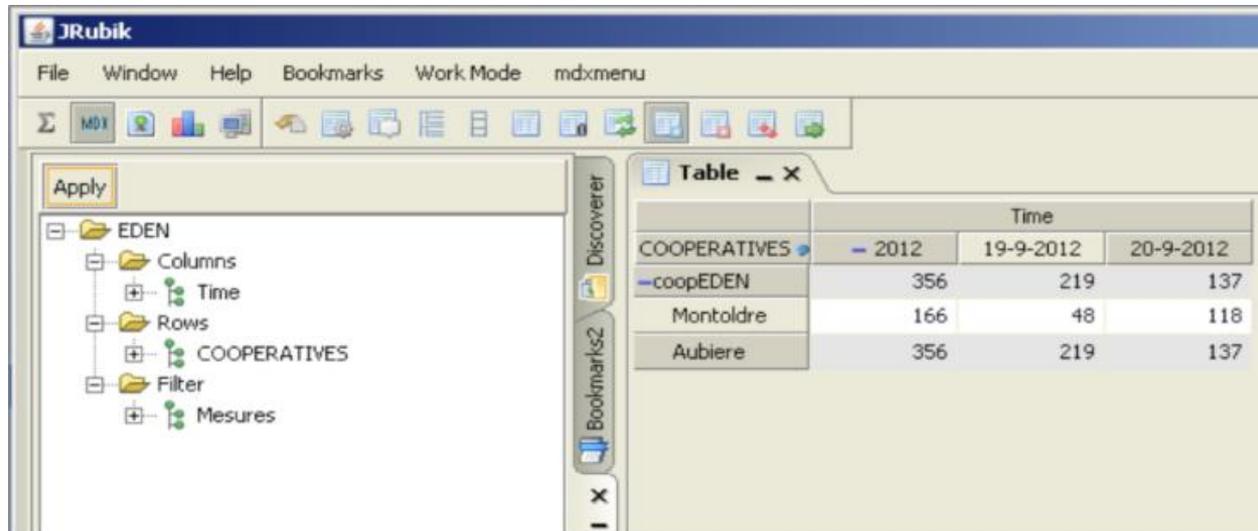
Feed dimensions Define measures Feed Fact Table Delete

Measure min value 0 Measure max value 254

ProtOLAP tool: Analysis Tier

La visualisation du cube prototype peut se faire avec tout client qui supporte XMLA (service web pour serveurs OLAP)

Exemple avec JRubik



The screenshot shows the JRubik application window. The menu bar includes File, Window, Help, Bookmarks, Work Mode, and mdx:menu. The toolbar contains various icons for navigation and analysis. The main interface is divided into several panes:

- Apply**: A tree view showing the cube structure. The root is 'EDEN', which contains 'Columns', 'Rows', and 'Filter'. 'Columns' includes 'Time', 'COOPERATIVES', and 'Mesures'.
- Discoverer**: A vertical pane with 'Discoverer' and 'Bookmarks2' buttons.
- Table**: A data table showing the results of a query. The table has columns for 'COOPERATIVES' and 'Time' (with sub-columns for '2012', '19-9-2012', and '20-9-2012').

COOPERATIVES	Time		
	2012	19-9-2012	20-9-2012
coopEDEN	356	219	137
Montoldre	166	48	118
Aubiere	356	219	137



ProtOLAP at work

ProtOLAP a été utilisé dans 3 projets concrets

Différentes publications scientifiques

Expérimentations :

L'usage de ProtOLAP diminue considérablement le temps de conception des cubes OLAP

- de quasi 5 fois



ProtOLAP demain ?

ICSOLAP peut être implémenté dans d'autres CASE tool (ex: Eclipse) à la place de MagicDraw

Le support pour d'autres SGBD peut être implémenté

Finalisation de l'outil WIKI

ProtOLAP demain ?

Finalisation de ProtOLAP & Transfert industriel

....





Thanks!

Questions?

Sandro Bimonte

Sandro.bimonte@irstea.fr

IRSTEA, Clermont-Ferrand, France



www.irstea.fr

