

Cycle de conférences autour du SI : Glossaire BI

Cycle 1 : le SI Décisionnel - Mardi 2 octobre 2012

1 Généralités

Qu'est-ce que l'informatique décisionnelle (Business Intelligence) ?

Ce sont

1. **les moyens, les outils et les méthodes** qui permettent de
2. **collecter, consolider, modéliser, exploiter et restituer les données** d'une Université et de permettre à la
3. **Gouvernance (présidents, secrétaire généraux), contrôleurs de gestion mais aussi les services** d'avoir une
4. **vue métiers ou d'ensemble, une vision cohérente et compréhensible** de l' Université, à des fins de
5. **statistiques, d'analyse, de contrôle ou de pilotage** de l'activité traitée en vue d'offrir une ...
6. **Aide à la décision**

1.1 Indicateurs (KPI's Key Performance Indicators)

Un indicateur est un outil d'évaluation et d'aide à la décision (pilotage, ajustements et rétro correction) grâce auquel on va pouvoir mesurer une situation ou une tendance, de façon relativement objective, à un instant donné, ou dans le temps et/ou l'espace. Un indicateur se veut être une sorte de résumé d'informations complexes offrant la possibilité à des acteurs différents (gouvernance, gestionnaires, etc.) de dialoguer entre eux. L'indicateur (qualitatif ou quantitatif) décrit généralement un état, ne pouvant être appréhendés directement. Un indicateur peut en agréger d'autres. L'utilité d'un indicateur dépend d'abord de sa capacité à refléter la réalité, mais aussi de sa simplicité d'acquisition et de compréhension.

Les indicateurs offrent un moyen d'évaluer les progrès en vue d'un objectif. Ex : Nombre d'étudiants décrocheurs au 1er semestre L1

1.2 Objectifs

Ensemble de résultats à atteindre issue de la stratégie d'un établissement que l'on suit à l'aide d'indicateurs. Ex :

- Objectif : Renforcer l'attractivité de chercheurs / Indicateur : Proportion de chercheurs étrangers
- Objectif : Adapter la gestion financière et comptable / Indicateur : Capacité d'autofinancement

1.3 Analyse

L'analyse décisionnelle permet à des décideurs d'aller plus loin que la simple description des comportements passés et de prédire l'avenir. Elle permet de déceler dans les comportements des règles qui les induisent. Grâce à ces découvertes, il devient possible en université :

- De renforcer l'attractivité et la réussite
- De réduire les coûts
- D'optimiser la gestion de l'établissement

1.4 Compte Rendu (Reporting)

Désigne une famille d'outils destinés à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format prédéterminé sur des supports sous forme d'indicateurs. Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement.

1.5 Outil de restitution

Ensemble des outils (requêteurs, tableaux de bord, etc.) permettant de restituer le résultat d'une analyse.

1.6 Tableau de bord (Dashboard)

Echantillon réduit d'indicateurs permettant à un gestionnaire de suivre l'évolution des résultats, les écarts par rapport à des valeurs de référence (objectifs fixés...), en se concentrant sur ceux qu'il considère comme les plus significatifs.

1.7 Tableau de bord prospectif (scorecard)

Le tableau de bord prospectif (TBP) ou tableau de bord équilibré (en anglais, Balanced Scorecard ou BSC), est une méthode lancée en 1992 par Robert S. Kaplan et David Norton visant à mesurer les activités d'une entreprise en quatre perspectives principales : apprentissage, processus, clients et finances.

1.8 Processus de travail (workflow)

Un workflow (anglicisme) est la représentation d'une suite de tâches ou opérations effectuées par une personne, un groupe de personnes, un organisme, etc. Le terme flow renvoie au passage du produit, du document, de l'information, etc., d'une étape à l'autre.

2 Usages

2.1 Datamining (Exploration de données)

Le Datamining est un terme générique englobant toute une famille d'outils facilitant l'exploration et l'analyse des données contenues au sein d'une base décisionnelle de type DataWarehouse ou DataMart. Il permet de mieux comprendre les liens entre des phénomènes en apparence distincts et d'anticiper des tendances encore peu discernables. Avec le datamining on essaye d'établir des corrélations entre des données afin d'en tirer des renseignements, des indicateurs, des anomalies, des correspondances, etc. qui peuvent mettre en évidence des tendances. Celle-ci permet de constater qu'un fait, par exemple le taux d'encadrement, et de le décliner comme taux d'encadrement par composante, de classer ce fait, de l'éclairer en révélant par exemple les variables ou paramètres qui pourraient faire comprendre pourquoi ce taux d'encadrement de telle composante est supérieur à celui de telle autre.

2.2 Drill down (Forage)

Mécanisme de navigation dans une structure multidimensionnelle permettant d'aller du plus global au plus détaillé à l'aide de requêtes de type hiérarchique. C'est la possibilité de « zoomer » sur une dimension (par exemple d'éclater les années en 4 trimestres pour avoir une vision plus fine, ou de passer du pays aux différentes régions),

2.3 Axe d'analyse (ou Dimension)

Axe d'analyse est l'expression d'un contexte (*la dimension*) dans la mesure (*le fait*), exemple :

« Je souhaite connaître mon mes recettes (dont conventions) en ressources propres, subventions, ressources internes, prestations internes par Centre Financier, par structure analytique, par UFR et sur 4 ans»

C'est donc une approche « multidimensionnelle ».

2.4 Analyse multidimensionnelle

Concept qui définit les analyses effectuées par croisement de minimum trois dimensions (ou ensemble de données du même type ou encore axes).

2.5 SIO (Système d'information opérationnel)

Représente aujourd'hui la majeure partie du SI focalisé sur le fonctionnement de l'applicatif et du stockage des données opérationnelles avec des contraintes immédiates de niveau et de qualité de service, ainsi que des procédures répétitives, des transactions élémentaires, des données détaillées abondantes et peu lisibles, avec une absence de synthèses, une absence de sémantique commune, pas de recul historique, un contenu très riche et de faible valeur informationnelle. (*voir aussi OLTP*)

2.6 SID (Système d'information Décisionnel)

Destiné uniquement à produire de l'information et non à automatiser des opérations, est découplé du SIO mais alimenté par le SIO. Il transforme les données pour améliorer leur valeur informationnelle, est potentiellement concerné par tous les types de données. Les informations sont des « données mises en forme », la valeur d'une information ne dépend pas uniquement de celle des données sous-jacentes, l'information n'est pas bonne ou mauvaise, elle est simplement appropriée ou non, le SID transforme des données brutes en informations appropriées. (**voir aussi OLAP**)

3 Technique

3.1 Infocentre

Un infocentre consistait dans les années 70 et 80 à mettre à la disposition d'utilisateurs finaux toute la puissance de calcul d'un ordinateur en temps partagé au moyen de terminaux, de banques de données, de langages

L'infocentre en tant que base de données mise à disposition des utilisateurs a, depuis les années 90, été remplacé par l'informatique décisionnelle, et les concepts de DataWarehouse, de DataMart ou entrepôt de données.

3.2 BdD (Base de données)

Lot d'informations stockées dans un dispositif informatique. Typiquement elle renferme toute l'information d'un domaine métier (BdD Apogée - > Formation, BdD Sifac - > Finance ...).

3.3 ETL (Extract Transform Loading.)

Outil chargé d'extraire, nettoyer et transformer les données émanant de sources diverses pour ensuite les insérer dans une base de données cible (DataWarehouse, DataMart, etc.).

3.4 DataWarehouse (ou entrepôt de données).

Un DataWarehouse est une base de données, spécifique au monde décisionnel, utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles multiples et fournir une aide à la décision à l'Université. Un Entrepôt de données est une base de données regroupant l'ensemble des données fonctionnelles d'une entreprise. Il entre dans le cadre de l'informatique décisionnelle ; son but est de fournir un ensemble de données servant de référence unique, utilisée pour la prise de décisions dans l'entreprise par le biais de statistiques et de rapports réalisés via des outils de.

3.5 DataMart (Entrepôt de données intelligent)

Un DataMart (magasin de données) est un sous-ensemble d'un DataWarehouse destiné à fournir des données aux utilisateurs, et souvent spécialisé vers un groupe, métier ou un type d'actions.

Techniquement, c'est une base de données relationnelle utilisée en informatique décisionnelle pour restituer des informations ciblées sur un métier spécifique, constituant pour ce dernier un ensemble d'indicateurs utilisés pour le pilotage de l'activité et l'aide à la décision. Dans un DataMart les données sont ciblées, organisées, regroupées et agrégées pour répondre à un besoin spécifique à un métier ou un domaine donné. Il est donc destiné à être interrogé sur un panel de données restreint à son domaine fonctionnel, selon des paramètres qui auront été définis à l'avance lors de sa conception.

3.6 OLTP (On Line Transaction Processing)

Modèle utilisé par les SGBD opérationnels (issus des SIO). Son mode d'exploitation est transactionnel. L'objectif est de pouvoir insérer, modifier et interroger rapidement les données de la base de données. Ainsi, OLTP est caractérisé par un grand nombre de transactions de type INSERT, UPDATE, DELETE. L'accent est mis sur le traitement des requêtes très rapides, le maintien de l'intégrité des données dans des environnements multi-accès et une efficacité mesurée par le nombre de transactions par seconde. Dans les bases de données OLTP, les données sont détaillées et à jour.

3.7 OLAP (On Line Analytical Processing)

Modèle utilisé par les SGBD décisionnels (issus des SID). Caractérisé par un volume relativement petit de transactions, mais sur un volume très important de données. Les requêtes sont très complexes et impliquent de nombreuses agrégations. Pour les systèmes OLAP, la donnée temps de réponse n'est pas très importante. Ce système

travaille en lecture seulement. Les programmes consultent d'importantes quantités de données pour procéder à des analyses. Les objectifs principaux sont : regrouper, organiser les informations provenant des sources diverses, les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier, retrouver et analyser l'information facilement et rapidement. Cela nécessite de consulter des versions historisées de la base et peut se permettre d'ignorer temporairement les dernières mises à jour. Ces bases sont souvent d'un ordre de grandeur nettement supérieur à celles des bases OLTP, du fait de la conservation de l'historique.

OLPT	OLAP
Les utilisateurs	
<p>Les utilisateurs SONT les « rouages » de l'université</p> <p>Beaucoup d'utilisateurs concurrents</p> <p>Ne traitent qu' 'un seul domaine métier à la fois</p> <p>Exécutent un grand nombre de fois la même tâche</p> <p>Lisent et modifient les données en grand nombre</p> <p>Exigent des performances temps réel</p>	<p>Les utilisateurs OBSERVENT les « rouages» de l'université</p> <p>Peu d'utilisateurs</p> <p>Traient beaucoup de domaine métier simultanément</p> <p>Exécutent peu de tâche</p> <p>Lisent uniquement les données qui sont souvent récapitulées</p> <p>Performance des outils de restitution et d'analyse nécessaire</p>
Le contenu des données	
<p>Nécessaire au fonctionnement de l'université</p> <p>Normalisé</p> <p>Non historisé</p> <p>Mise à jour – Maintenant (now)</p>	<p>Utilisée pour analyser le fonctionnement de l'université</p> <p>Dénormalisée souvent agrégées</p> <p>Historisée (le temps est fondamental)</p> <p>Observation : Passé, Projection voire simulation</p>

3.8 ODS (et Staging area)

Operating Data Store ou magasin de données opérationnelles permet de stocker les données extraites des SI sources et de faire des opérations sur ces données. Dans le cas du staging Area, les données sont détruites directement après avoir été chargées dans le Datawarehouse mais pas pour l'ODS où les données ont une durée de vie plus longue. ODS répond plus à une problématique de reporting immédiat dans le sens où l'ODS sera mis à jour plus souvent que le Datawarehouse : on pourrait dire que l'ODS pourrait être alimenté tous les jours / semaine et le Datawarehouse une fois par mois. l'ODS n'est donc pas forcément indispensable si l'université peut recharger son Datawarehouse à fréquence < mois

4 Concepts

4.1 Fait

Un fait désigne les données observables d'une analyse menée dans le cadre de l'aide à la prise de décision que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyse (les dimensions). Par exemple : taux d'encadrement, nb d'inscrits, budget constituent des faits. Les « faits », dans un entrepôt de données, sont normalement numériques, puisque d'ordre quantitatif.

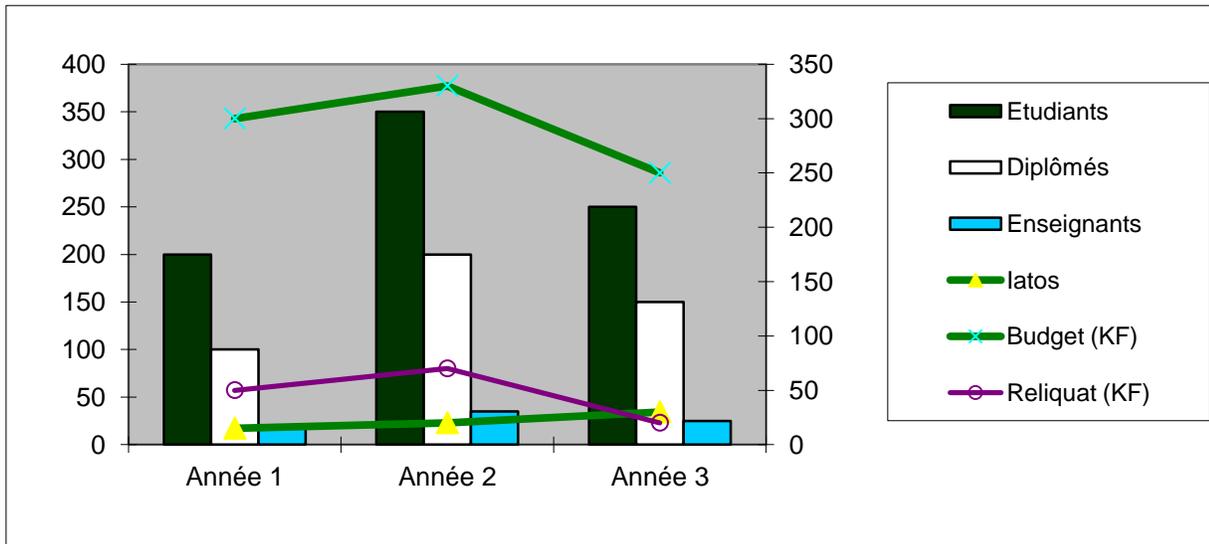
4.2 Table de Faits

Les faits doivent être liés à des tables, qui constituent alors des tables de fait. Chaque fois qu'on définit une table comme table de fait, un fait du même nom que la table est automatiquement créé.

4.3 Dimension (ou axe).

Une dimension est une table qui contient les axes d'analyse selon lesquels on veut étudier des données observables (les faits) qui, soumises à une analyse multidimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision.

On appelle donc « dimension » un axe d'analyse. Il peut s'agir des fournisseurs, d'une période de temps comme un exercice financier, ou calendrier scolaire (dimension asynchrone), d'une structure d'une UFR etc



Exemple: quelques indicateurs :

- Nb d'étudiant,
- De diplômés,
- d'enseignants,
- de personnel IATOS,
- de dépense budgétaire et
- du reliquat

pour une composante donnée.

L'année est une dimension (Axe 1) , la composante pourrait être une seconde dimension (Axe 2), l'entité géographique (campus) une troisième (Axe 3).

4.4 Cube

Structure multidimensionnelle permettant l'analyse d'informations factuelles en les segmentant sur un ensemble d'axes d'analyses.

4.5 Métadonnées

Informations qui renseignent sur la nature de certaines données. Les métadonnées que l'on peut par exemple associer à un document sont : son titre, son auteur, sa date de création, etc. Dans le cadre du décisionnel, elles constituent une sorte de dictionnaire sur lequel le système s'appuie pour comprendre des données utilisées par les différentes applications qui alimentent le DataWarehouse. Les intitulés "Fournisseurs" d'un PGI finance comme SIFAC et "nom" d'une application comptable peuvent contenir les mêmes informations mais le système ne peut le savoir que si un dictionnaire a été conçu pour lui indiquer qu'il s'agit de la même nature d'informations. Les métadonnées englobent

également l'ensemble des informations relatives à la provenance, à l'historique et aux traitements associés aux données d'un DataWarehouse.

4.6 Couche sémantique

C'est une représentation métier des données, elle offre une interface simple à utiliser et compréhensible permettant à des utilisateurs non spécialistes des logiciels de, d'exécuter des requêtes sur une base de données afin d'effectuer des analyses et de créer des rapports.

Par exemple la couche sémantique sous BO est appelée Univers est composé de 2 types de structures : les classes, regroupement logique d'objets de même thème dans un univers, et les objets l'élément le plus fin de l'univers, il représente des données de la base.

4.7 In Memory

High-Performance Analytic Appliance (HANA). C'est le fait de créer une organisation des données décisionnelles, appuyée a de la technologie associative (vectorielle) et de l'analyser en mémoire. Dédié aux services opérationnels et marketing pour permettre de travailler de manière plus efficace