

Business Intelligence, Etat de l'art et perspectives

ICAM – JP Gouigoux – 10/2012



CONTEXTE DE LA BI

Un peu d'histoire

- Premières bases de données utilisées comme simple système de persistance du contenu des applications
- Besoin de reporting pour les managers, ainsi que pour les analystes, pour qui les données de gestion représentent des statistiques intéressantes (sources d'économie, de redirection ou de confirmation de stratégie, de lancement de campagnes)

Tableaux de bord en 2001 (V6)

MGDIS - Observatoire

MGDIS [Fichier | Edition | Affichage | Format | Données | Outils | Compléments | Mylis] [Consultation] [100%]

Fichier | Edition | Affichage | Format | Données | Outils | Compléments | Mylis [R-Dossiers (1)]

Etat du dossier	Dossier - Code	Dossier	Service du gestionnaire
TOTAL			
Clôturé	IDF_ACTCO_13221_P_00 00000012	Mettre en place mener et mesurer le programme d'action Visite SOPEMEA IDF GS78 14/05/2008	GS Yvelines
Commencé	ALS_ENT_1009_S_00 ALS_ENT_1070_S_00 ALS_ENT_1084_S_00 ALS_ENT_1090_S_00 ALS_ENT_1091_S_00 ALS_ENT_1096_S_00 ALS_ENT_1099_S_00 ALS_ENT_1104_S_00 ALS_ENT_1105_S_00 ALS_ENT_1106_S_00 ALS_ENT_111_S_00 ALS_ENT_1113_S_00 ALS_ENT_1123_S_00 ALS_ENT_1140_S_00 ALS_ENT_1148_S_00 ALS_ENT_1156_S_00 ALS_ENT_1166_S_00 ALS_ENT_117_S_00 ALS_ENT_1172_S_00 ALS_ENT_1180_S_00 ALS_ENT_1181_S_00 ALS_ENT_119_S_00 ALS_ENT_12_S_00 ALS_ENT_1206_S_00 ALS_ENT_1209_S_00 ALS_ENT_1212_S_00 ALS_ENT_122_S_00 ALS_ENT_1222_S_00 ALS_ENT_1225_S_00 ALS_ENT_1230_S_00 ALS_ENT_1245_S_00 ALS_ENT_125_S_00 ALS_ENT_1258_S_00 ALS_ENT_1268_S_00 ALS_ENT_1290_S_00 ALS_ENT_13_S_00 ALS_ENT_133_S_00 ALS_ENT_1363_S_00 ALS_ENT_1364_S_00 ALS_ENT_1372_S_00 ALS_ENT_1381_S_00 ALS_ENT_1393_S_00 ALS_ENT_140_S_00 ALS_ENT_1400_S_00 ALS_ENT_1401_S_00	TECHNI SOUDURE HUG successeur HOCHSTICHER Sarl NORMALU SUFRA MECASEM CEDAM TECHNIQUES METALLIQUES APPLIQUEES VOIRIN CONSULTANTS MARMPOSE EST XYNATECH EUROPE GEHOLIT VIVRAD PAUL MATHIS SA APPLICATIONS TECHNOLOGIES NOUVELLES ACTIVIS BRICKMANN STOCKMEIER URETHANES FRANCE SCHROFF SA TARACELL FRANCE LARISYS ESCALIERS BECK BEI-DEACOD LIEBHERR FRANCE SA PISCINES WATERAIR BOOMERANG Pharmaceutical Communications EMI MEI MANUFACTURE D IMPRESSION SUR ETOFFES SEMPATAP CONSEIL ET HOMMES ATIKA SYSTEM AMBOS HORUS XTREM NUMERIQUE GENERALE DE POTERIES D ALSACE IREPA LASER GAGGENAU Industrie VOSSLON-SCHWABE FRANCE INCOTEC SATIE SA NATURUNION AEM ACEAN SCIERIES NOLLINGER DE FIL EN AIGUILLE IMPREX	

En savoir plus... [Définir]

Informations disponibles

- R-Dossiers
 - Données générales
 - Domaine du métier analysé
 - Référence
 - Objet du dossier
 - Dossier père - Code
 - Dossier père
 - Tiers - Code
 - Tiers
 - Etat
 - Date de création
 - Date de modification
 - Modifié par - Code
 - Modifié par
 - Date de début d'instruction
 - Date de clôture
 - Etat du dossier - Code
 - Gestionnaire
 - Direction du gestionnaire
 - Direction du gestionnaire - Référence
 - Gestionnaire - Code
 - Gestionnaire
 - Procédure - Code
 - Procédure
 - Profil - Code
 - Profil
 - Profil - Procédure - Code
 - Profil - Procédure
 - Code du groupe de directions
 - Libellé du groupe de directions
 - Code de l'unité
 - Libellé de l'unité
 - Référence de l'unité

Thématiques

Graphiques et cartes associées

Anal [Fichier | Edition | Affichage | Format | Données | Outils | Compléments | Mylis] [Consultation] [100%] [Aucun calcul] SOGOS 11:57

Montée en puissance de la BI

- Volumes élevés, granularité faible : il faut agréger
- Approche semi-automatique (« simili-OLAP »)

Tableaux de bord en 2009 (WEB)

MG005 - PROGOS.NET

Tableaux ... (2)

POP4- Détail limitrophe- Pop 15 ans ou + par sexe, âge et vie en couple -- RRPOP4_DETGeo-1

Navigation : Consultation

Fichier Affichage Données

Exercice (*)2007

Niveau géographique (*)Commune

Commune limitrophe (*)ARRADON

Libellé géographique	Sexe	Âge regroupé (7 classes d'âges 15-80+ ans)	Population de 15 ans et plus	Population de 15 ans et plus (%)	dont: Vie en couple	% Vie en couple	dont: Ne vivant pas en couple	% Ne vivant pas en couple
Total			52 762	100.0%	27 395	51.9%	25 367	48.1%
BADEN			3 228	100.0%	2 292	71.0%	936	29.0%
	Homme		1 577	46.9%	1 145	72.8%	432	27.4%
		15 - 19 ans	116	3.6%	3	2.6%	113	97.4%
		20 - 24 ans	79	2.4%	6	7.6%	73	92.4%
		25 - 39 ans	277	8.6%	197	71.1%	80	28.9%
		40 - 54 ans	447	13.8%	378	84.6%	69	15.4%
		55 - 64 ans	301	9.3%	262	87.0%	39	13.0%
		65 - 79 ans	293	9.1%	255	87.0%	38	13.0%
		80 ans ou plus	64	2.0%	44	68.8%	20	31.3%
	Femme		1 651	41.1%	1 147	69.5%	504	30.5%
		15 - 19 ans	101	3.1%	5	5.0%	96	95.0%
		20 - 24 ans	52	1.6%	14	26.9%	38	73.1%
		25 - 39 ans	278	8.6%	227	81.7%	51	18.3%
		40 - 54 ans	468	14.5%	385	82.3%	83	17.7%
		55 - 64 ans	332	10.3%	290	87.3%	42	12.7%
		65 - 79 ans	303	9.4%	198	65.3%	105	34.7%
		80 ans ou plus	117	3.6%	28	23.9%	89	76.1%
PLOEREN			4 511	100.0%	3 091	71.7%	1 420	28.3%
	Homme		2 315	49.3%	1 546	79.3%	569	26.9%
		15 - 19 ans	195	4.5%	0	0.0%	195	100.0%
		20 - 24 ans	119	2.8%	13	10.7%	106	89.3%
		25 - 39 ans	511	11.9%	406	79.3%	106	20.7%
		40 - 54 ans	720	16.7%	640	89.0%	79	11.0%
		55 - 64 ans	274	6.4%	246	89.6%	28	10.4%
		65 - 79 ans	270	6.3%	224	82.7%	47	17.3%
		80 ans ou plus	25	0.6%	17	71.0%	7	29.0%
	Femme		2 196	50.7%	1 545	70.4%	651	29.6%
		15 - 19 ans	185	4.3%	3	1.6%	182	98.2%
		20 - 24 ans	94	2.2%	30	32.1%	64	67.9%
		25 - 39 ans	593	13.7%	492	83.1%	100	16.9%
		40 - 54 ans	694	16.1%	603	86.9%	91	13.1%
		55 - 64 ans	330	7.7%	261	79.0%	69	21.0%

Disponibles

Données définies sur la consultation

Informations hiérarchiques

Informations calculées

Informations de structure

Population de 15 ans ou plus par sexe, âge et vie en couple

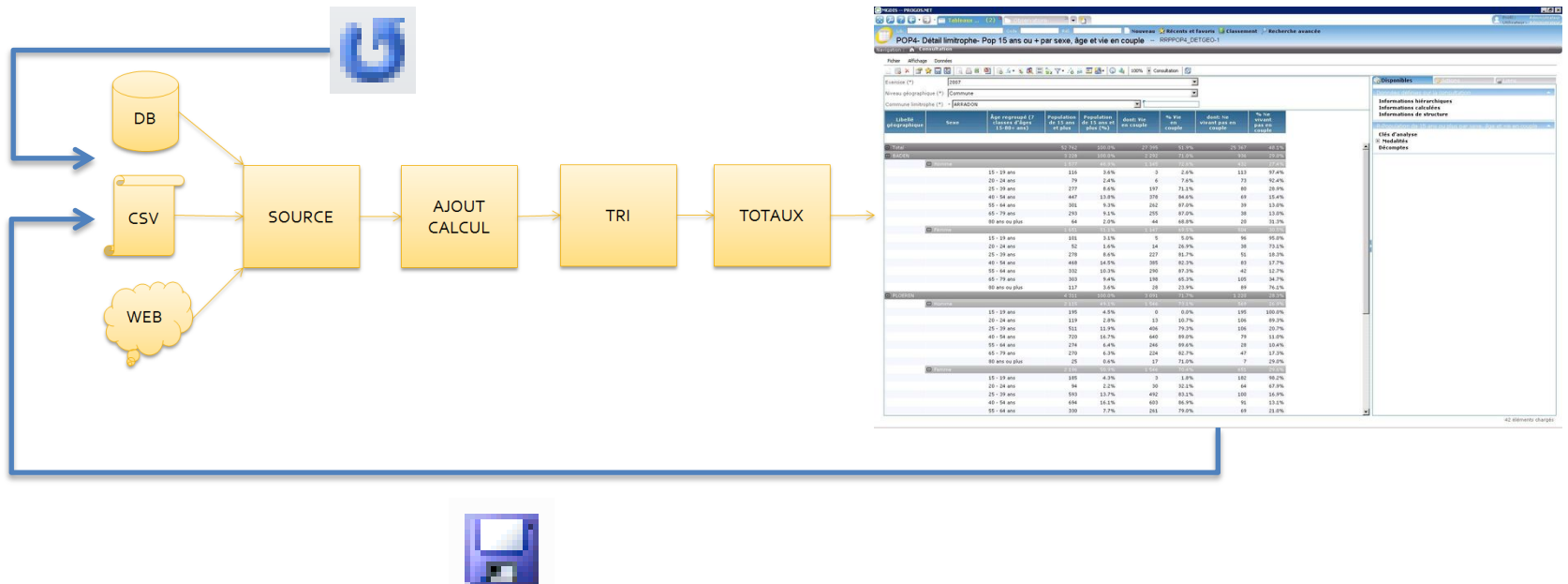
Clés d'analyse

Modalités

Décomptes

42 éléments chargés

Agrégation semi-automatique



OLTP / OLAP

- Transaction / Analytique
- OLTP : optimisation pour l'écriture et la mise à jour, avec recalcul rapide de quelques indexes
- OLAP : optimisation pour la lecture, avec de nombreux indexes, quitte à avoir un recalcul de ces derniers sur période creuse
- Séparation faits (mesures) / dimensions (clés)

Cubes OLAP

- Pré-agrégation de toutes les combinaisons de dimensions
- Exemple : date agrégée par jour, semaine, mois, trimestre, semestre, année
- Sql Server Analysis Services
- Pentaho

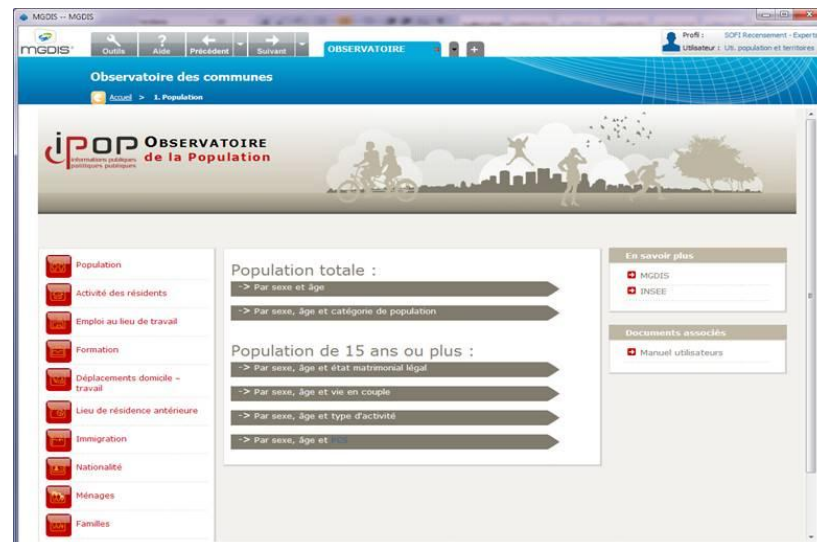
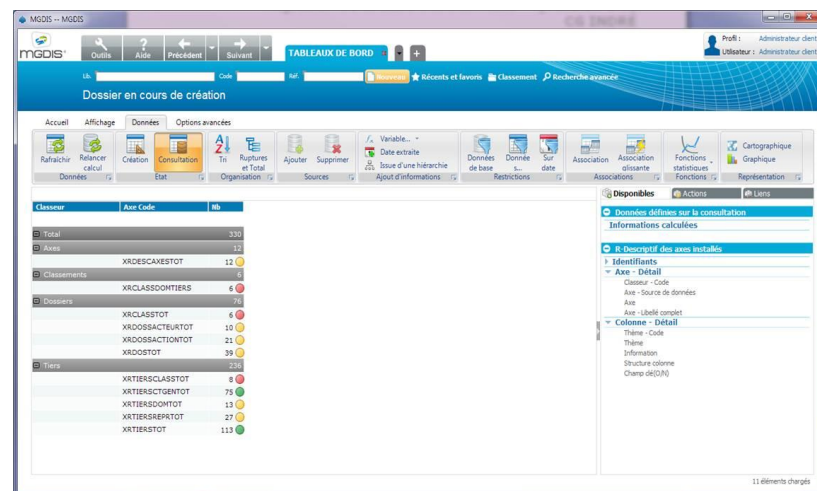
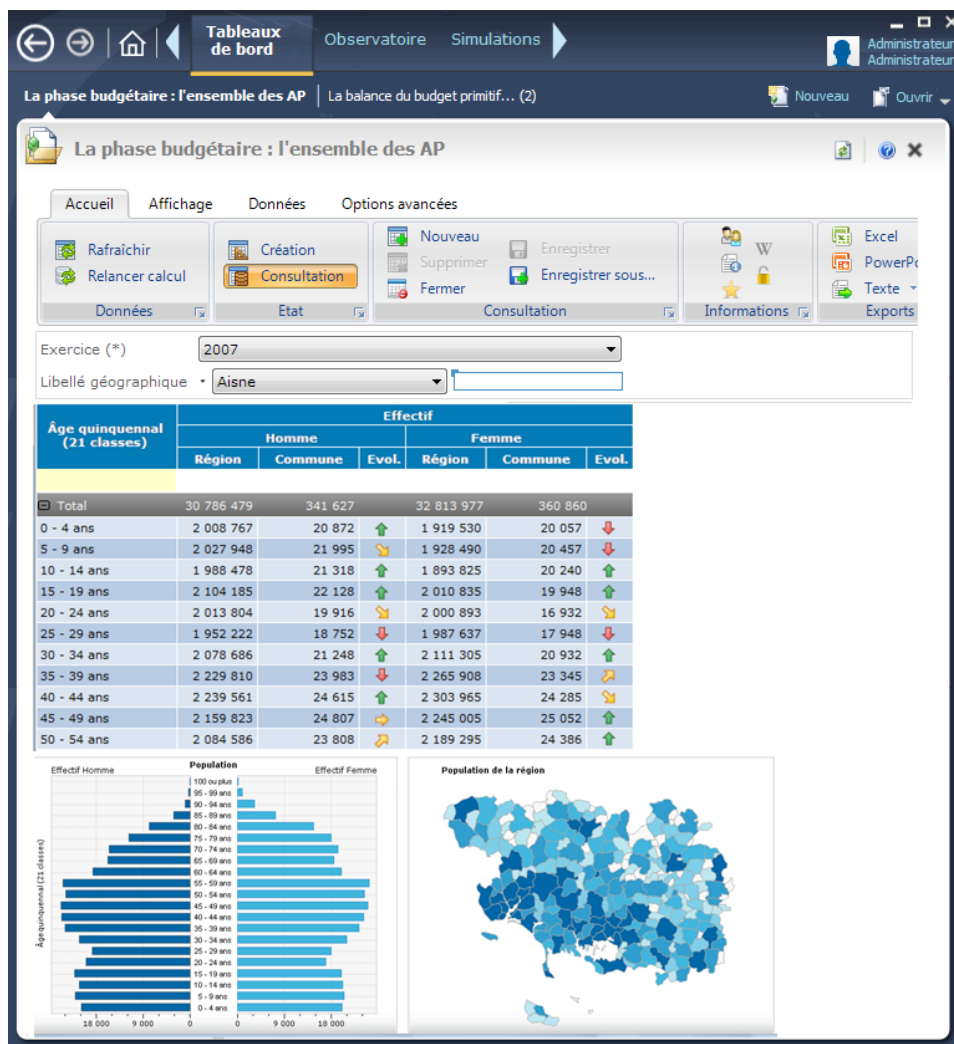
Traitements de données

- Extract / Transform / Load
- Sql Server Integration Services
- Talend
- Temps réel / Temps différé

Data warehousing

- Séparation asynchrone
- Datamart / data warehouse
- Modélisation en étoile
 - Mesures (faits) au centre
 - Dimensions autour, du plus au moins granulaire
- Business Objects

Tableaux de bord en 2012 (SPADE)





ÉTAT DE L'ART : I HAVE BAD NEWS

BI is dead

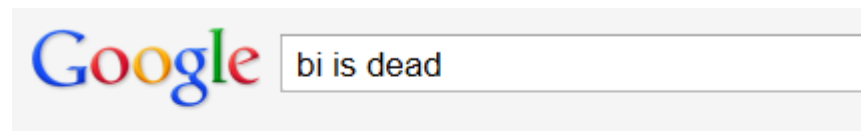
- Chute libre pour Gartner

CIO Technologies		Ranking of technologies CIOs selected as one of their top five priorities in 2009.			
Ranking		2010	2009	2008	2007
Virtualization		1	↑ 3	3	5
Cloud Computing		2	↑ 16	*	*
Web 2.0		3	↑ 15	15	*
Networking, voice and data communications		4	↑ 6	7	4
Business intelligence (BI)		5	↓ 1	1	1
Mobile Technologies		6	↑ 12	12	11
Data & Document Management and Storage		7	↑ 10	9	9
Service-oriented applications and architecture		8	↑ 9	10	7
Security technologies		9	↓ 8	5	6
IT Management		10	* *	*	*
Enterprise Applications		11	↓ 2	2	2

* New Question

BI is dead

- Tom Gonzales : is BI broken ?



Recherche

Environ 429 000 000 résultats (0,11 secondes)

because I believe too much of the focus is still being p
users interact with the data, and what business condi

Does the BI Industry have it all wrong?

Big BI (defined by the small group of large but influen
difficult to escape. As a result, they are slowing down

had consciously started to focus my attention on the dashboard side of the business. After engaging in many well intentioned and executed BI projects that yielded mediocre results at best, I started to wonder "Is BI broken? Is there a better way?" At the time, I saw a focus on dashboards solving many

was being sold as the sizzle as opposed to the steak? This same conundrum holds true for most of the industry experts v
developing and evangelizing rigorous disciplines and techniques to build these data Taj Mahals. What do they do now? S
-- all that stuff we told you to do for the past 15 years is really not the most effective approach--we got it backwards."

ointing the way to the future of BI?

BI is dead

- Taux d'échec des projets pire qu'en IT

Les tendances IT du Gartner jusqu'en 2016 ! - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://www.lemondedcloud.fr/lire-les-tendances-it-du-gartner-jusqu-en-2016-46879-page-2.html

Google

Les tendances IT du Gartner jusqu'e...

Mise à jour le 09/12/2011 17:42

Les sites ITNewsInfo : CMC RT Dis

Le Monde Informatique en partenariat avec ORACLE

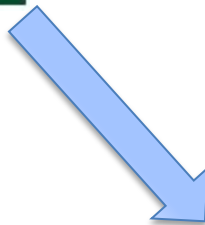
Livre blanc Datacenter

- ❑ Gamme de solutions Dell pour la virtualisation du poste de travail
Découvrez toutes la gamme de solutions de Dell Virtualisation avec les modèles de prestation, les av...
- ❑ Optimiser la virtualisation intégral du poste de travail
La virtualisation du poste de travail permet à chacun de travailler partout, à toute heure, souven...
- ❑ Mobilité et virtualisation du poste de travail
Mobilité, appareils personnels variés, applications multiples, réseaux omniprésents, tout concourt ...
- ❑ Cinq conseils pour des sauvegardes et restaurations efficaces dans les environnements

Données transférées depuis fpdownload2.macromedia.com

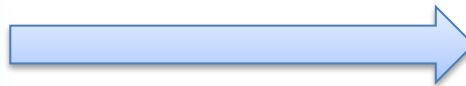
Echec de l'approche

- Projets menés par la DSI
 - Technique
 - Centralisation



Echec sur l'utilisation

- Questions métier
 - Comment je me compare ?
 - D'où vient le problème ?
 - Toujours des tableurs...



[REFUSÉ]

Echec du marché

- Oligopole
 - Software : Microsoft, Oracle, BO
 - Hardware : IBM, HP
 - Course à l'armement
- Open Source arrivé trop tard



Echec intellectuel

Customer Challenges

- Increased volumes of data
- Need to reduce costs
- Increased adoption of DW appliances
- Move to MPP
- Demand for flexibility and mixed workloads across the enterprise
- Desire for real-time analytics
- Growing importance of data quality

Approximate data volume managed by data warehouse

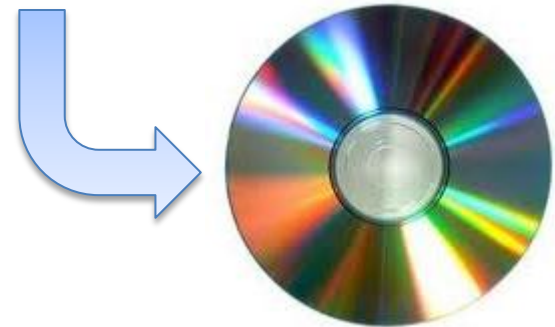


More than 10 TB

17%

34%

Microsoft



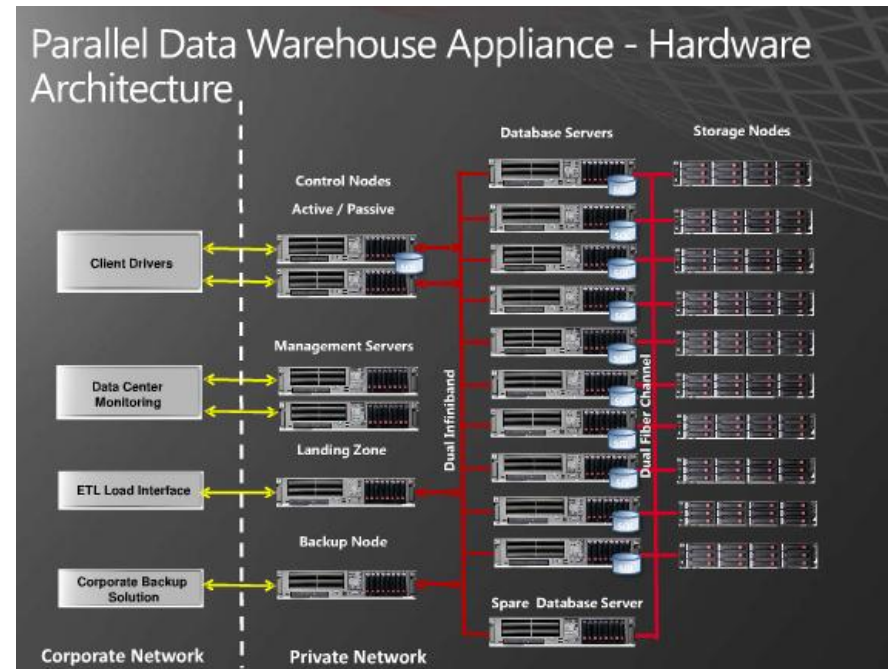
<http://www.slideshare.net/msitpro/sql-server-parallel-data-warehouse>

"We don't have better algorithms than anyone else. We just have more data."

--Peter Norvig, Chief Scientist, Google

Echec de l'innovation

- Toujours plus
 - RAC, Parallel DW
 - Clusters
 - Exadata, Hana
- Limites
 - Types structurés
 - Affinité
 - Centralisation (SVOT)
 - Redondance



<http://www.slideshare.net/msitpro/sql-server-parallel-data-warehouse>

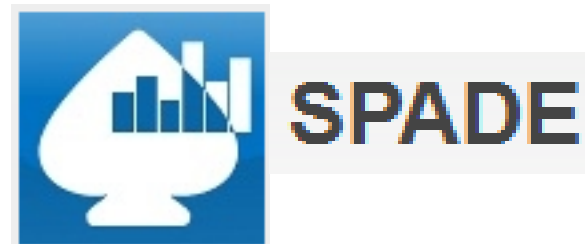
DES SOLUTIONS ?

Approches légères

- Utilisabilité supérieure à la consistance SVOT



QlikView



PowerPivot (sources)

Classeur1 - Microsoft Excel

PowerPivot pour Excel - Classeur1

Accueil Conception

Type de données : Nombre entier
Format : Général

Tableau croisé dynamique

Obtenir des données externes

N° tour	Code département	Code de la commune	Nom de la commune	N° de bureau de vote	Inscrits	V
2	62	1	Ablain-Saint-Nazaire	1	727	
2	62	2	Ablainzeville	1	151	
2	62	3	Acheville	1	463	
2	62	4	Achicourt	1	839	
2	62	5	Achiet-le-Grand	1	767	
2	62	6	Achiet-le-Petit	1	280	
2	62	7	Acq	1	392	
2	62	8	Acquin-Westbécourt	1	504	
2	62	9	Adinfer	1	174	
2	62	10	Affringues	1	183	
2	62	11	Agnez-lès-Duisans	1	567	
2	62	12	Agnières	1	159	
2	62	13	Aigny	1	1662	
2	62	14	Aire-sur-la-Lys	1	1062	
2	62	15	Airon-Notre-Dame	1	167	
2	62	16	Airon-Saint-Vaast	1	180	

PR07_T2-Bvot

Enregistrement : 1 sur 131 234

Assistant Importation de table

Connexion à une source de données
Vous pouvez créer une connexion à une source de données ou utiliser une connexion existante.

Bases de données relationnelles

- Microsoft SQL Server**
Créez une connexion à une base de données SQL Server. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Microsoft SQL Azure**
Créez une connexion à une base de données SQL Azure. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Microsoft SQL Server Parallel Data Warehouse**
Créez une connexion à un entrepôt de données SQL Server Parallel Data Warehouse. Importez des tables ou des vues, ou des données retournées par une requête.
- Microsoft Access**
Créez une connexion à une base de données Microsoft Access. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Oracle**
Créez une connexion à une base de données Oracle. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Teradata**
Créez une connexion à une base de données Teradata. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Sybase**
Créez une connexion à une base de données Sybase. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Informix**
Créez une connexion à une base de données Informix. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- IBM DB2**
Créez une connexion à une base de données DB2. Importez des tables ou des vues de la base de données, ou des données retournées par une requête.
- Autres (OLEDB/ODBC)**
Créez une connexion à une source de données en utilisant un fournisseur OLEDB ou ODBC pour ODBC. Importez des données à partir des tables ou des vues retournées...

Sources multidimensionnelles

- Microsoft Analysis Services**
Créez une connexion à un cube SQL Server Analysis Services. Importez les données retournées par une requête MDX.

Flux de données

- Rapport**
Créez une connexion à un rapport Microsoft Reporting Services. Importez les données de la source.
- Dataset Azure DataMarket**
Créez une connexion à un dataset Azure DataMarket. Importez des données du dataset.
- Autres flux**
Créez une connexion à un flux de données. Importez les données du flux.

Fichiers texte

- Fichier Excel**
Importez des données d'un fichier Excel.
- Fichier texte**
Importez des données à partir d'un fichier texte.

< Précédent Suivant > Terminer Annuler

PowerPivot (TCD)

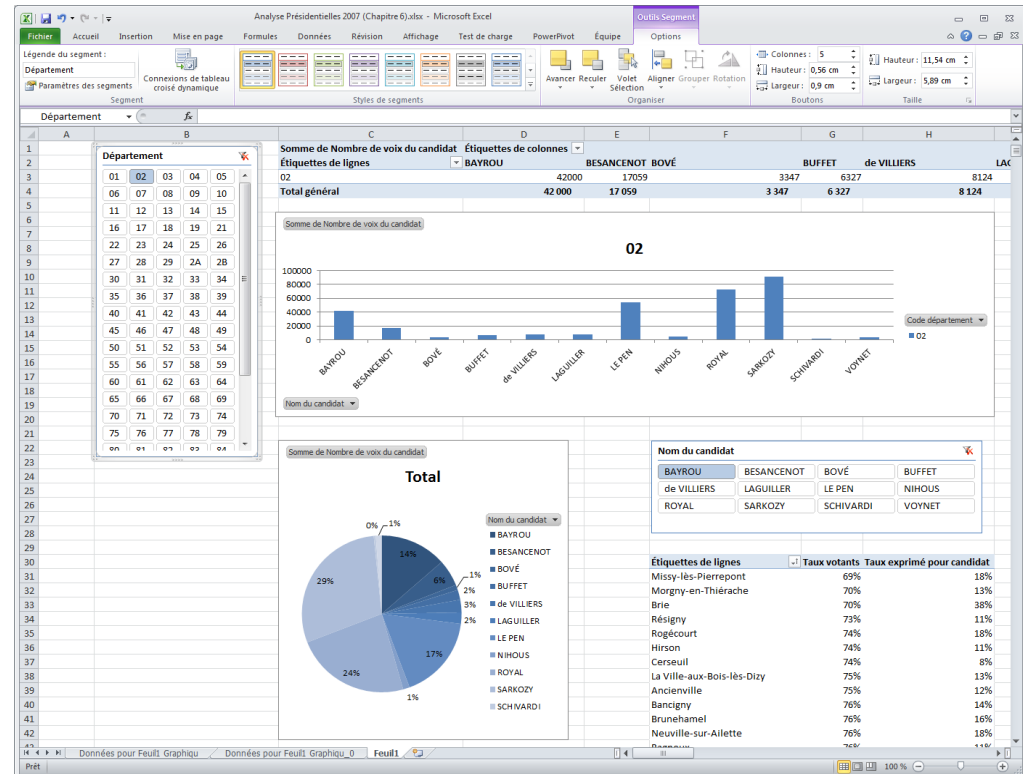
Liste de champs PowerPivot

Choisissez les champs à ajouter au rapport :

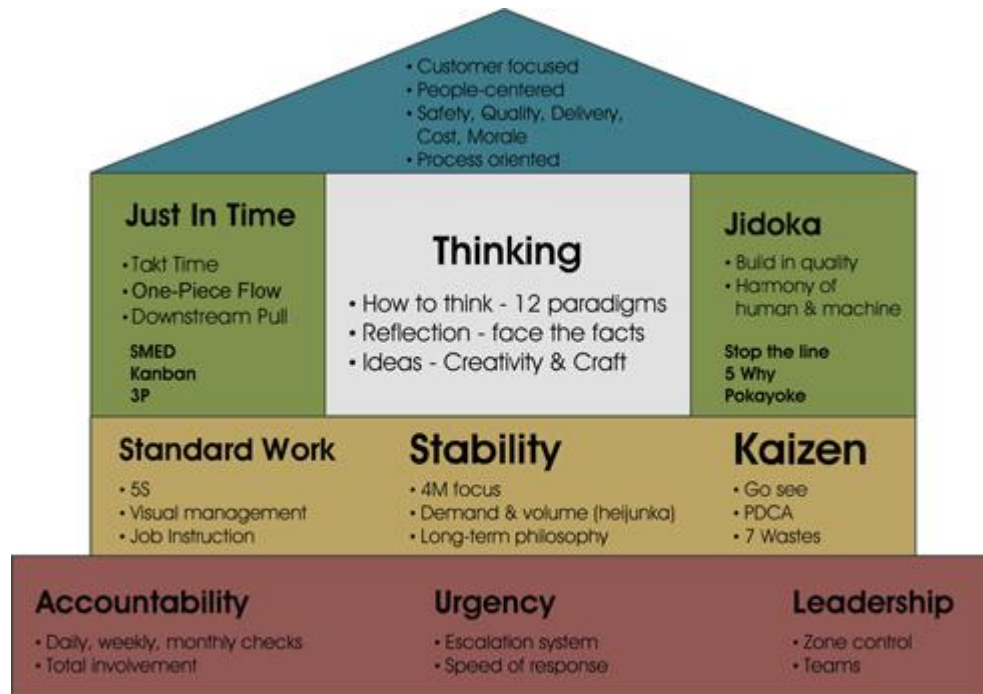
Rechercher

- ☒ PR07_T2-Bvot
 - ☒ N° tour
 - ☒ Code département
 - ☐ Code de la commune
 - ☒ Nom de la commune
 - ☐ N° de bureau de vote
 - ☐ Inscrits
 - ☐ Votants
 - ☐ Exprimés
 - ☐ N° de dépôt du candidat
 - ☒ Nom du candidat
 - ☐ Prénom du candidat
 - ☐ Code nuance du candidat
 - ☒ Nombre de voix du candidat

Somme de Nombre de voix du candidat		Étiquettes de colonnes	
Étiquettes de lignes	ROYAL	SARKOZY	Total général
Brest	40217	30504	70721
Lannion	7269	4113	11382
Rennes	60538	35995	96533
Vannes	13193	14790	27983
Total général	121217	85402	206619



Think Lean



Taiichi Ohno

Lean BI

- Lean = traquer le gaspillage
- Pagination
- Curseur coût / gain
- Optimisation logicielle
- Changement architecture

Agilité dans la BI ?



- Développement
 - 40% de projets en échec
 - Manifeste agile pour repartir du client
- Idem pour BI
 - Repasser en bottom-up (Respect)
 - Abandonner la centralisation (Pragmatisme)
 - Abandonner le SVOT (Courage)

APPROCHE DISTRIBUEE

Utilisation du Cloud

- Scalabilité
- Elasticité des coûts
- Disponibilité
- Multi-tenancy

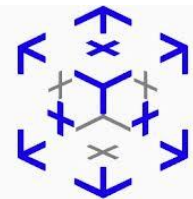


Les limites du Cloud

- Psychologique
- Politique
 - Patriot Act
 - Safe Harbour Act
 - Cloud européen
 - Cloud français
 - Andromède
 - Numergy / CloudWatt
- Sécurité des données
 - Contractuelle
 - Blocage propriétaire



Cloudwatt™



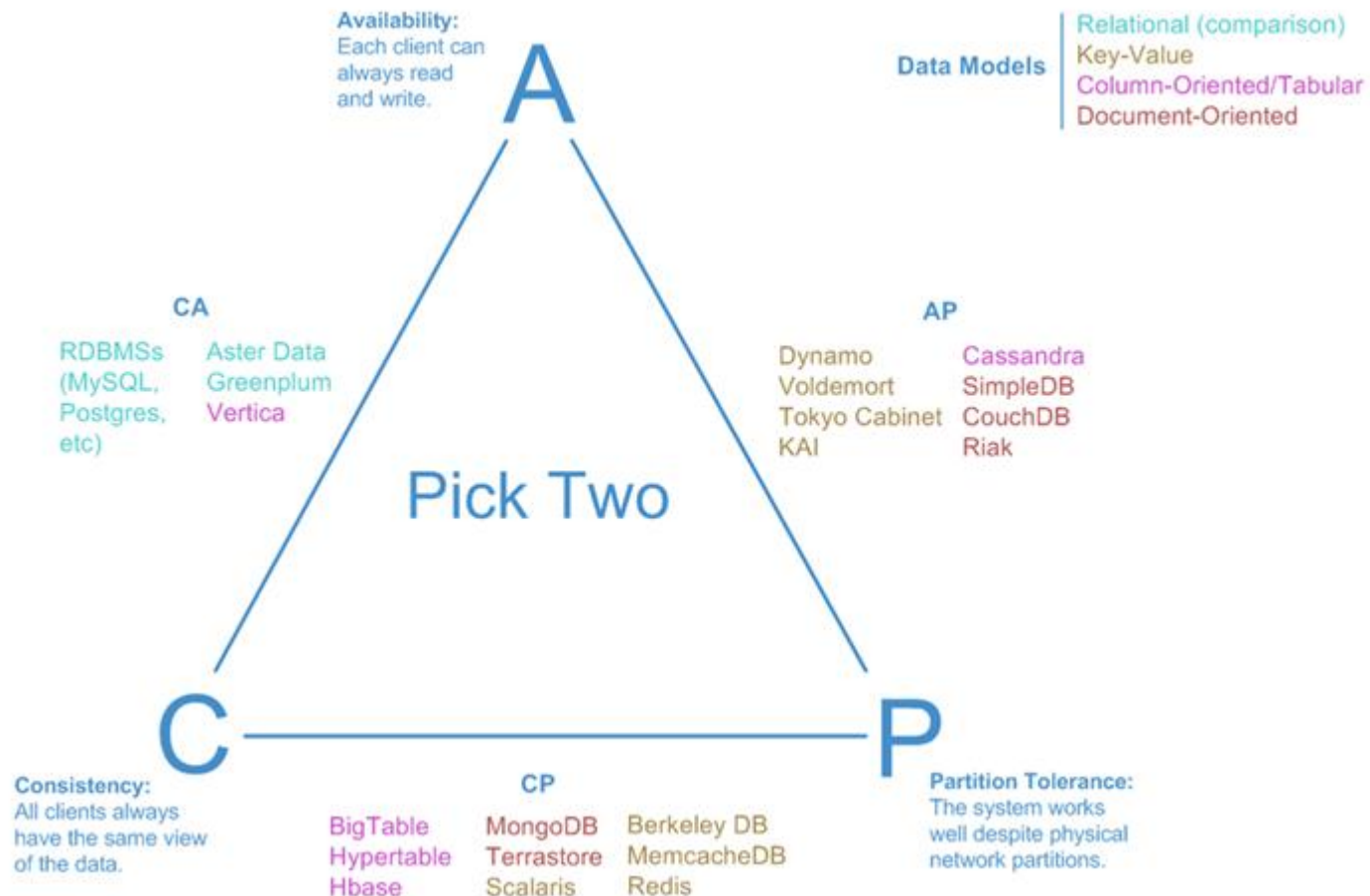
NUMERGY

Approche NoSQL

- Modèle tabulaire trop rigide
- BASE au lieu d'ACID
- Requête ad-hoc

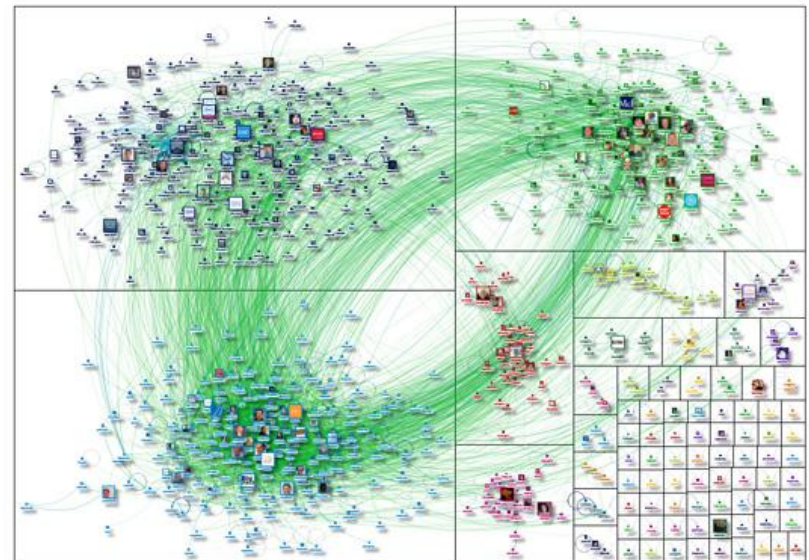


Triangle NoSQL



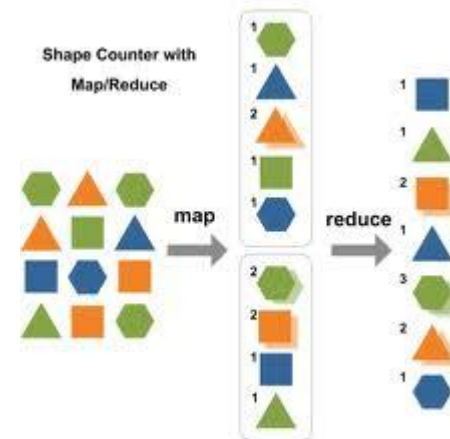
Approche « Big Data »

- Trois V
 - Volume (video-surveillance)
 - Variety (logs serveurs)
 - Velocity (API Twitter / Facebook)
- Plus tard (révélateur)
 - Value



Bases de Big Data

- Map Reduce
 - Algorithme de découpage
 - Algorithme de calcul distribué
 - Algorithme de recomposition
- HDFS
- Hadoop



Usages de Big Data

- Gestion de logs
- Détection de signaux faibles
 - Intrusion réseau
 - Comportements sur réseaux sociaux



OPEN DATA

Aspect politique

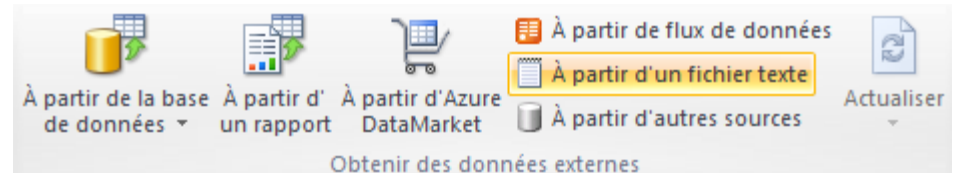
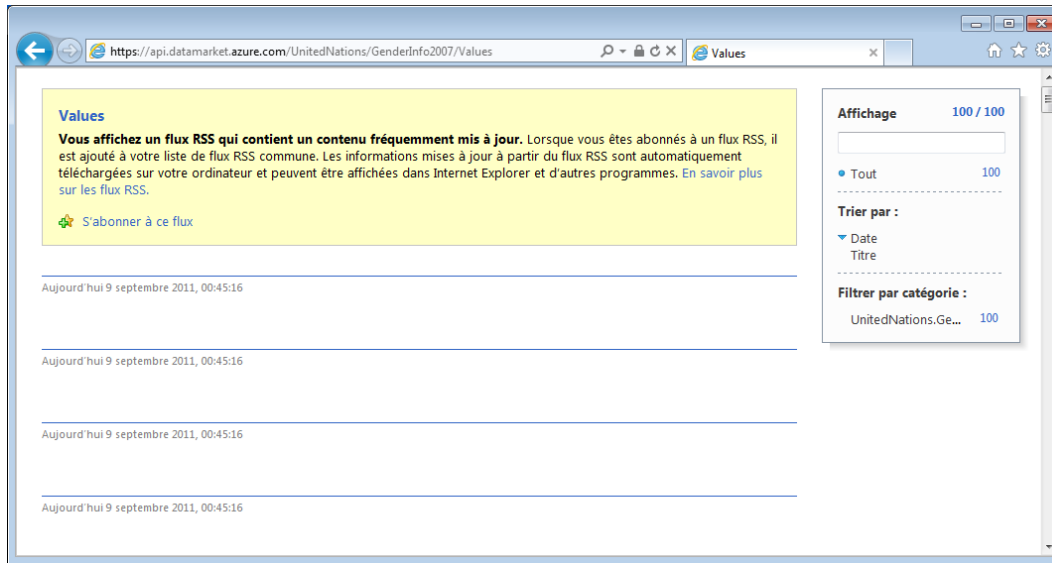
- data.gov (2007)
- data.gov.uk (2009)
- ETALAB
- data.gouv.fr (2012)
- Collectivités locales en avance
 - Rennes
 - CUB
 - CG71



Normalisation de la donnée

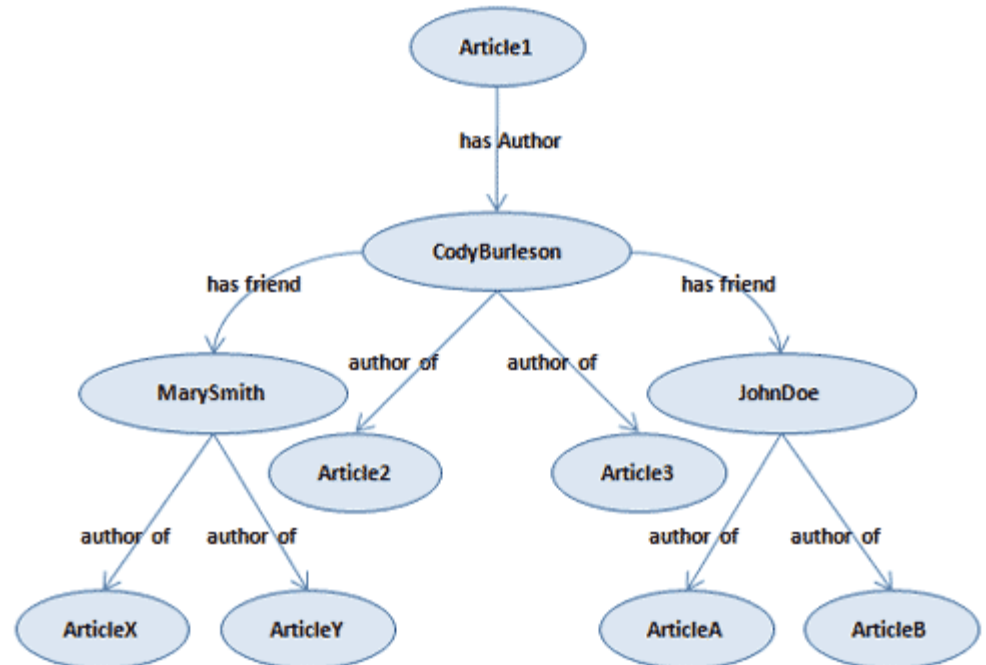
- data.gouv.fr : 352 000 sources, 90% CSV
- Utilisabilité obérée
- Solution par la norme
 - GData
 - OData
 - Protocole en V2
 - APIs et outils disponibles pour nombreux systèmes

Protocole OData



Etape suivante : les métadonnées

- Triplette RDF
- Ontologie OWL
- data.gov.uk en avance



Limites d'Open Data

- Données décentralisées, mais pas les relations
- Jointures techniques
- Manque de métadonnées de base
 - Unités
 - Normes utilisées (exemple de la population)
- Diffusion anarchique

Opportunités Open Data

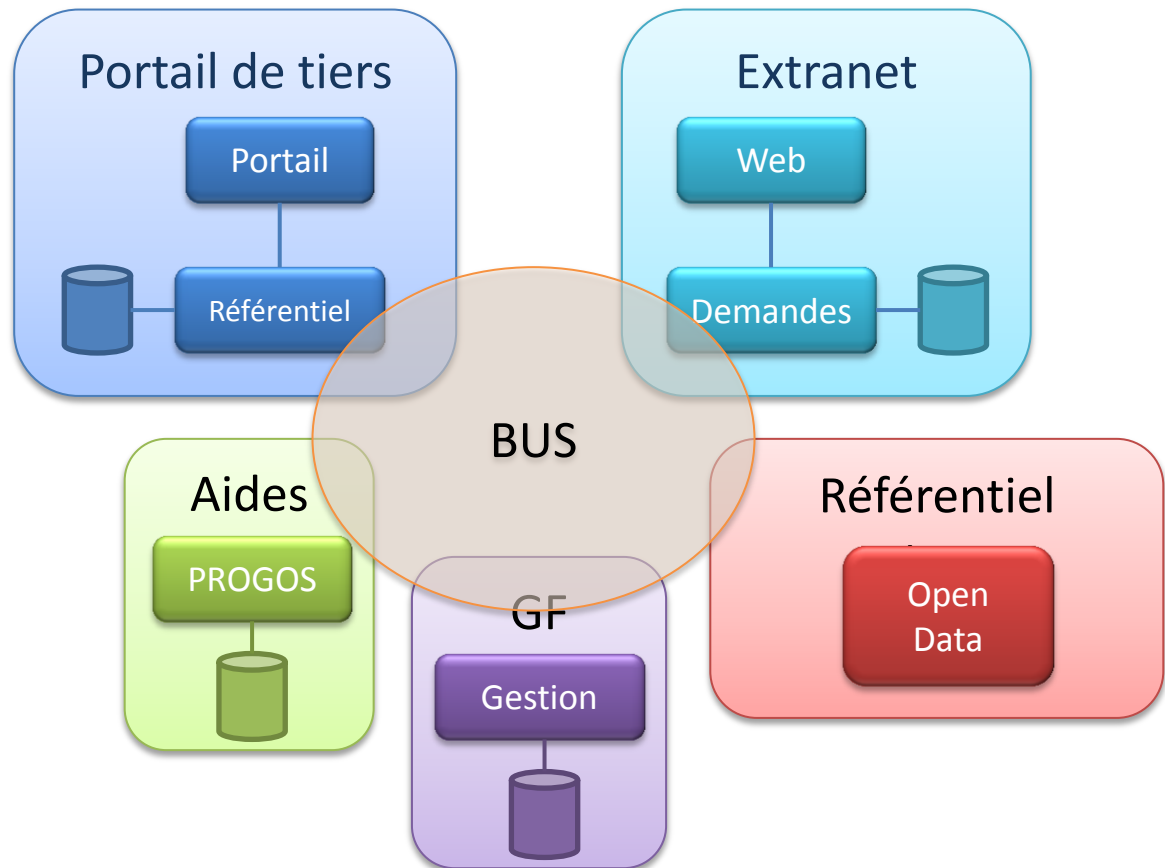
- Normalisation
- Economie de ressources
- Réduction des erreurs
- Impact citoyen



INTEGRATION DANS UN ESB

Concepts de SI en ESB

- Découplage
 - Routage
 - Médiation
 - Garantie de livraison



Master Data Management

- Données communes
- Routage
- Cache de données extérieures

BI dans un ESB

- Approche côte-à-côte
 - Couche services
 - Couche données
 - ETL sur les bases de données
- Pattern de service de données
 - Lève le problème de couplage
 - Versioning progressif

LE FUTUR DE LA BI

Veille technologique (2010)

- 28/11/2010
- « BI 4, l'ultime suite décisionnelle de SAP »

« BI 4 offre le moyen d'embrasser la totalité des données de l'entreprise sans aucune restriction, avec une couche sémantique évoluée et une intégration avec les données multi-dimensionnelles indispensables au pilotage », précise Laurent Leenhardt, le directeur des solutions métiers de SAP France. Selon lui, la BI est devenue tout terrain. Comprenez qu'elle se consomme partout et de différentes manières, depuis un terminal mobile, une tablette, via un widget, via des portails, au travers d'outils de collaboration, en push, en pull, via des applications, sur papier...

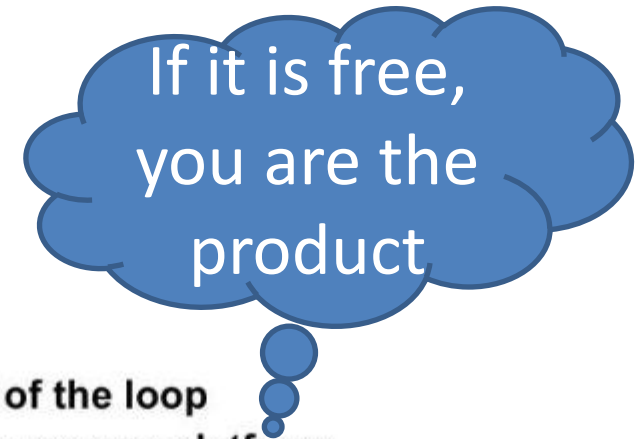
L'acquisition de Business Objects a réellement transformé le visage de SAP.

La nouvelle suite d'aide à la décision prolonge la gamme de progiciels de SAP par des outils collaboratifs structurants. Ils aident à propager de bonnes pratiques au bon moment. En amont, l'intégration des données est facilitée. Il s'agit de croiser les données de la production, des ressources humaines et d'autres services de l'entreprise pour offrir un portail décisionnel fiable et sûr et optimiser les performances de chaque activité.

Avec le modèle "on-demand", SAP entend offrir un accès immédiat, à la demande, aux meilleures pratiques de gestion. Cette externalisation des moyens d'analyse peut intervenir en complément des applications décisionnelles d'entreprise, pour soutenir une surcapacité saisonnière par exemple. Troisième axe, le déploiement 'on-premise' permet aux entreprises de continuer à se développer autour de serveurs d'applications sur site.

Veille technologique (2011)

- Données hétérogènes
 - Mashup
 - Collaboratif
- Mobilité
 - Localisation
 - Cartographie
 - Voice (Siri)
 - QRCode
- Interfaces
 - Touch
 - Réalité augmentée

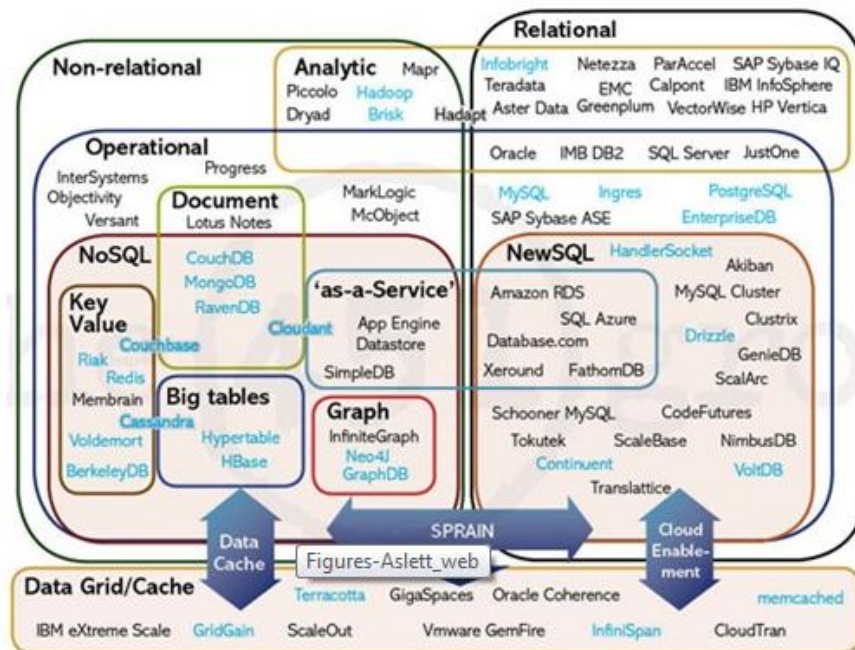


- The PC is out of the loop
- The phone is a sensor platform
 - Hardware add-on innovation
 - Location based sensing
 - Touch screen UI
- Processing is done in real time in the cloud
 - Allowing processing that can't be done on the device
 - Big data analysis
 - Building new networks on the back of existing ones
- Reinventing a major industry

<http://www.slideshare.net/timoreilly/the-future-of-business-intelligence>

Veille technologique (2012)

- Arrêter de rêver
- Retrousser les manches
- Viser la BI comme commodité
- NewSQL



CONCLUSION

Si vous ne vous rappelez que ceci

- Plenty of room at the bottom
 - Partir des besoins de l'analyste
 - Utiliser des architectures modernes
 - Réaliser, tester, reboucler
- Il y aura toujours un invariant dans le système
 - La BI aide la réflexion
 - Elle ne la remplace pas