1 Consultants

Olivier Decoq et Flavien (il était à l'université de Mons).

2 BUSINESS intelliJ

Pannel d'application et de technologies afin de les analyser, stocker et les mettre à disposition par la suite. Optimisation des noix pour l'écureuil. Différent niveau d'analyse, reporting que c 'est t il passé, analysis pourquoi ça c 'est passé comme ça monitoring que c'est il passé dans le systeme prediction que va t il se passer dans le systeme.
Predictive analysis : combien de client vont pouvoir venir visiter mon magasin la semaine prochaine ? Tourné vers le futur, quelle est la tendance projetée par rapport aux deux prochaines années. Le futur, le déjà vu : prédictive analysis.
Process BI collecter des données, les analyser, en faire de la connaissance qui a du sens ; Peremet de prendre des décisions et prévoir des actions qui vont définir le business.
Présentation du gateau.
Rendre les données SEXY et exploitable, puis manger le gateau.

3 Data WAREHOUSE

On prends l'excell on leur dit que c'est 10 000 à la place de 100 et tout le monde croit que c'est 10 000. But : être sur que c'est la vérité. Il y a différentes sources hétérogènes ex base de données oracke, sql, ... fichiers plats, fichiers textes, excell, xml,... On remplit de cette façon un data warehouse. A partir de ça, on crée des techniques d'accès à la demande avec des alertes etc... Façon plus académique : ensemble de données en voyant ses sujets intégré, historisé,..
Intégré : nettoyer des données pour les mettre dans le data warehouse exemple : on va refaire un encodage pour avoir un encodage unique (x,y) = (Masculin, Féminin)
Historisation : 5 à 10 ans les données voir + en fonction des besoins des personnes qui les utilisent. A tout moment on doit etre capable de recréer l'info qui étaient dans le data warehouse. Pour se faire : structure en charpente dans le data warehouse ; notion de temps .
Structuration en étoile. Approche : Bill Mole voit tout en top down il modélise d abord l ensemble du data warehouse par rapport a toute l entreprise etc marketing ressources humaines etc et il va créer un modèle complet en ayant une vue complète de la société à l'ancienne. Une fois qu'il aurta tout modélisé il va développer l'implémentation physique. Inconvénient : bcp de temps pour l'analyse. Avantage : vision globale de la société et moins de travail sur les différentes données car on aura déjà pensé a tout. Kimball Cho : Un sujet de métier à la fois. Il se base sur un sujet. Il fait les sujets métiers 1 à 1 et considère que l'ensemble de son datamarkt constitue son data warehouse. Il va falloir fussionner pour pouvoir faire transiter l'info d'un sujet à l'autre. Plus y a de data markt, plus y aura de travail à faire. Quand on en a 4 5 6, facile mais 20 = spaghetti. Gros avantage : on peut fournir à notre business des données sur lesquelles il peut jouer.

Tous les jours si on veut mettre des nouvelles données dans le data wharouse, ça prend plusieurs heures, ça dépend du volume. Exemple assurances, 30 Millions d euros par jour => table plus grosse, en terme d'exploitation, plus difficile. Data wharehouse de plusieurs TERA (le fameux jeux vidéo). Plus il y a de données à charger, plus ça prend du temps + les utilisateurs demandent d'avoir accès à leur données in real time ; problématique qui permet d'introduire la partie de flavien : BIG DATA.

4 BIG DATA : la digitalisation

Les données sont absolument partout : twitter facebook, recherches google, image
No respect de la vie privée.. Il se sert de monticules de données pour savoir les gouts de la personne.
"Big data = grosse données"
-Jakie Chan
Start up californiene : GOOGLE. Le but : pouvoir indexer toute sles pages de l'internet et pouvoir tenir compte des changements qu'il y avaient sur ces pages internet. Représente des milliards de pages. Volume de données colossale.
Challenge dont google a fait face : tous les outils qui existaient à l'époque étaient dépassés.
3 axes : volume de données, la vélocité (tenir compte de toutes les màj sur les sites en temps réels), la variété (sites web avec textes, images, vidéos ou encore que sais je > Stocker image dans base de données pas facile). Plutot que de se baser sur une base de données et de rajouter de la mémoire, débit, .. Ils se sont reposés sur une infrastructure en cluster càd plusieurs machines reliées à un data center
Le google filesystem ne stocke pas un fichier volumineux sur une seule mchine et divise ce fichiers en blocs qui sont dispatchées sur une machine du cluster puis répliqué ! Quand on se retrouve avec de smilliers de machines interconnectées et la défaillance devient la règle > pleins de répliques. Big table : BDD distribuée pas stockée sous forme de table mais plutot sous forme de clé valeur ; Si on veut stocker les informations d'un site internet, ex doctissimo, avec des infos sur les médicaments etc. Si on veut stocker ces infos dans une BDD classique ; collonne Soin de sante, médicaments,... COmment deviner le choix de la table ? Google va pouvoir stocker avec leur systeme n'omprte quelle donnée.
Map reduce : façon de programmer ; Les fichiers ne sont pas stockés sur une machine mais répartis en bloc sur tout un tas de machine. + facile : distribuer le calcul sur l'ensemble de leur machine sur des plus petits fichiers.
Exemple : grand nombre de cartes de couleurs pour 10 personnes, on donne un sous ensemble de ces cartes et on demandent à chaque personne de les trier par couleur => clé valeur (couleur, nombre)/ ON réordonne tout cela par couleur.

5 Solution open source du meme genre : hadoop

Solution copier coller de google mais open source. (voir les slides)
Spark : comme map reduce mais in memory càd que à la fin du traitement

enregistre sur disque dur

6 Building your data science team

osef

7 USE CAAAAAAAAAAASE

Flu Trends ?
Prédire en temps réel la propagation de l'épidémie en fonction des recherches. / ! Pas pris en compte le feed back loope sous l effet des medias (ex médias effet bdn).

8 Les BSP

Fin du séminaire pour moi les gars