

1.BIG DATA

Big data regroupe ainsi des outils et techniques pour traiter les données qui répondent au défi du volume important, de leur variété en nature et en sources, d'un certain niveau de vitesse (on emploie plutôt le terme vélocité) et la véracité (traduit de l'anglais veracity, qui est à prendre dans le sens de vraisemblance).

Les 3 dimensions ou les 3V du Big Data :

- Le Volume
- La Variété
- La Vélocité

Le Volume :

Le premier des 3V Big Data concerne les volumes de ces données : ces volumes sont non seulement considérables mais ils sont en perpétuelle croissance. Aujourd'hui ces volumes s'expriment en téraoctets (10 puissance 12 octets, noté To). Et bientôt on utilisera plus couramment le pétaoctet (10 puissance 15 octets, noté Po).

La Variété :

Le 2ème des 3V Big Data se rapporte à leur variété : ces données sont en effet de formats très variés. Quoi de commun entre une vidéo, un morceau de musique, un message sur Twitter, une photo en ligne, un courrier papier, un post sur une page Facebook et un relevé de compteur électrique ou un relevé de banque par exemple. Les supports et les formats sont tous différents, mais ils portent tous une information de valeur.

La Vélocité :

Le 3ème des 3V Big Data est la vélocité des données : elle correspond à la vitesse à laquelle doivent être collectées, analysées et exploitées ces données. De plus en plus souvent, ces données sont à traiter en temps réel.

Deux autres dimensions se sont ajoutées par la suite à ces 3 dimensions essentielles de ces données (on parle aussi des « 5V du Big Data ») :

- La Véracité
- la Valeur

La Véracité :

La véracité ou validité des données correspond à la confiance que ces données inspirent aux décideurs. Si les utilisateurs de ces données doutent de leur qualité ou de leur pertinence, il devient difficile d'y investir davantage. Le nombre grandissant de fausses informations et de faux avis de consommateurs publiés sur le web est un bon exemple de données externes peu fiables dont les entreprises doivent pouvoir se prémunir.

La Valeur :

Cette dimension s'attache à un aspect étonnant de ces données. Une donnée prise isolément peut ne pas avoir grand intérêt, mais prise dans un ensemble plus large, elle prend tout son sens. C'est le cas par exemple d'un tweet portant sur un événement d'actualité. Le contenu du tweet en lui-même peut ne pas avoir grande importance, en revanche si le tweet est repris et commenté des milliers de fois, c'est le nombre de messages associés à ce tweet qui devient une information de valeur (une opinion partagée par plusieurs, un sujet qui fait consensus ou polémique...).

a) Avantages pour les entreprises :

Le ciblage comportemental prédictif

On peut savoir rapidement l'impact d'une campagne publicitaire et ainsi en déduire des analyses de performance des différentes campagnes. On peut en mesurer le succès contrairement aux autres formes de publicité.

b) Avantages pour les internautes :

La publicité en ligne comprend aussi quelques avantages pour les internautes voulant acheter un produit en particulier. En effet, on peut acheter plus rapidement et plus facilement en cliquant simplement sur la bannière publicitaire proposée. La publicité prédictive se base souvent sur les dernières pages Internet consultées.

Ainsi, les annonces seront plus pertinentes et utiles pour les consommateurs puisqu'elles sont ciblées selon les préférences des internautes et leurs habitudes d'achats.

2.CLOUD COMPUTING

Le cloud computing ou informatique en nuage est défini comme étant un système informatique qui fournit en tant que service une grande capacité de calcul. Le cloud computing se base sur des technologies disponibles sur Internet afin de fournir ce service à de multiples clients externes.

Le cloud computing est un service vendu à la demande. Ce service peut être facturé, au temps utilisé (l'heure, la minute, la seconde), à la bande passante utilisée ou encore au nombre d'utilisations du service. Ce service est ajustable, c'est à dire que l'utilisateur est libre de l'utiliser comme il le souhaite et à tout moment.

Il est totalement géré et administré par le fournisseur de service, permettant à l'utilisateur consommateur de n'avoir besoin que d'un accès à internet et d'un ordinateur.

Le cloud computing est l'évolution et la convergence de beaucoup de technologies. On compte parmi celles-ci le "grid computing", qui consiste en un ensemble de ressources hétérogènes permettant de lourds calculs, la virtualisation qui permet d'utiliser au mieux les ressources d'une machine ou d'un cluster de machines (VMWare, XenServer etc...). Ces technologies regroupent les "X as a Service" tel que le SaaS (Software as a Service), le PaaS (Platform as a Service) ou encore le IaaS (Infrastructure as a Service).

Les types :

A. Software as a Service (SaaS) :

Ce modèle de service est caractérisé par l'utilisation d'une application partagée qui fonctionne sur une infrastructure Cloud. L'utilisateur accède à l'application par le réseau au travers de divers types de terminaux (souvent via un navigateur web). L'administrateur de l'application ne gère pas et ne contrôle pas l'infrastructure sous-jacente (réseaux, serveurs, applications, stockage). Il ne contrôle pas les fonctions de l'application à l'exception d'un paramétrage de quelques fonctions utilisateurs limitées.

B. Platform as a Service (PaaS)

L'utilisateur a la possibilité de créer et de déployer sur une infrastructure Cloud PaaS ses propres applications en utilisant les langages et les outils du fournisseur. L'utilisateur ne gère pas ou ne contrôle pas l'infrastructure Cloud sous-jacente (réseaux, serveurs, stockage) mais l'utilisateur contrôle l'application déployée et sa configuration.

C. Infrastructure as a Service (IaaS) :

L'utilisateur loue des moyens de calcul et de stockage, des capacités réseau et d'autres ressources indispensables (partage de charge, pare-feu, cache). L'utilisateur a la possibilité de déployer n'importe quel type de logiciel incluant les systèmes d'exploitation.

L'utilisateur ne gère pas ou ne contrôle pas l'infrastructure Cloud sous-jacente mais il a le contrôle sur les systèmes d'exploitation, le stockage et les applications.

Ces trois modèles de service doivent être déployés sur des Infrastructures qui possèdent les cinq caractéristiques essentielles déjà citées.

Avantages et inconvénients du Cloud Computing:

Avantages :

La mise à disposition immédiate et le déploiement d'applications majeures et d'environnements de travail, avec une mise à jour des applications systématique, aussi, le client est exempté de la maintenance.

Le partage de données est relativement facile : chaque utilisateur du CC peut rendre disponibles ses données à d'autres utilisateurs. Il est possible de créer une plateforme virtuelle collaborative.

Un avantage qui ne concerne qu'un nombre assez réduit d'entreprises, mais qui n'en n'est pas moins l'un des plus importants du Cloud Computing : un calcul très puissant, ce qui permet la délocalisation des traitements bénéficiant ainsi de toutes les ressources et performances mises à disposition.

Le libre accès au client qui lui permet une connexion partout, évitant ainsi un VPN (réseau privé virtuel) dans l'entreprise.

L'accès est possible à partir d'un simple navigateur, il est possible d'être informé en continu de l'évolution de la plateforme du CC.

Les contrats reliant le fournisseur au client ne sont pas des engagements à long terme, ce qui offre une liberté totale. Les services sont facturés soit par abonnement mensuel, soit à la demande.

Coût : du fait que le même service est proposé à de nombreux utilisateurs, et que le service ne nécessite aucune installation matérielle (à part une connexion internet et un ordinateur), son coût en est nettement amoindri.

Le service est fiable puisque basé sur des infrastructures performantes tolérantes aux pannes.

Inconvénients :

La confidentialité et la sécurité sont les principales inquiétudes du système puisque l'hébergement se fait dans un environnement externe à l'entreprise. Le vol et/ou la mauvaise utilisation sont des risques considérables.

La connexion internet est une obligation.

3.DATA WAREHOUSE :

L'objet de l'entrepôt de données (data warehouse) est de définir et d'intégrer une architecture qui serve de fondation aux applications décisionnelles. Les entrepôts de données ou DW (Data Warehouse) sont de plus en plus utilisés pour exploiter et analyser une grande quantité de données et faciliter ainsi les processus de prises de décision. Ils servent aux systèmes informationnels d'une entreprise et ils permettent de sauvegarder de façon permanente et sans équivoque tout l'historique de l'entreprise. Cet historique en termes de données est mis sous certaines formes dans une base de données (BD) dite entrepôt, afin de répondre aux besoins d'affaire de la compagnie. Ces systèmes sont qualifiés par l'exploitation des données d'une entreprise afin d'aider à la prise de décision. L'objectif visé à travers

ce travail est d'aborder le thème de l'entrepôt des données en exposant son fonctionnement et sa mise en place, en déterminant ses objectifs. Mais il s'agit aussi de tenter de cerner ses enjeux et présenter une étude de cas permettant de mieux comprendre notre sujet.

« Le Data warehouse (entrepôt de données) est une collection de données orientées sujet, intégrées, non volatiles et historiées, organisées pour le support d'un processus d'aide à la décision » (Inmon, 94).

En effet, le data warehouse regroupe toutes les données fonctionnelles d'une organisation afin de délivrer un ensemble d'informations aux différents services d'une entreprise. L'objectif est d'extraire des données pertinentes à partir de bases de données de production et de les organiser suivant un modèle adapté afin de faciliter des prises de décision.

■ Données orientées sujet:

Le data warehouse est organisé autour des sujets majeurs de l'entreprise. Les données sont donc structurées par thèmes (ex: clients, produits, ventes, etc.). Il regroupe les informations des différents métiers.

■ Données intégrées :

Les données proviennent de plusieurs sources différentes. Avant d'être intégrées au sein du data warehouse elles doivent être mise en forme et unifiées afin d'en assurer la cohérence. Cela nécessite une forte normalisation, de bénéficier d'un référentiel unique et cohérent.

■ Données historiées :

Les données sont datées et persistent dans le temps .Chaque nouvelle données est insérée. Un référentiel de temps doit être mis en place afin de pouvoir identifier chaque donnée dans le temps.

■ Données non volatiles :

Un data warehouse veut conserver la traçabilité des informations et des décisions prises. Une fois insérées dans l'entrepôt, les données ne sont ni modifiées ni supprimées. Elles sont conservées pour des analyses futures.

Objectifs du data warehouse

L'important d'une entreprise réside dans les informations qu'elle possède. Les informations se présentent généralement sous deux formes : les systèmes opérationnels qui enregistrent les données et le Data Warehouse. En bref, les systèmes opérationnels représentent l'emplacement de saisie des données, et l'entrepôt de données l'emplacement de restitution.

Les objectifs fondamentaux du Data Warehouse sont :

a) Rendre accessibles les informations de l'entreprise : le contenu de l'entrepôt doit être compréhensible et l'utilisateur doit pouvoir y naviguer facilement et avec rapidité. Ces exigences n'ont ni frontières, ni limites. Des données compréhensibles sont pertinentes et clairement définies. Par données navigables, on n'entend que l'utilisateur identifie immédiatement à l'écran le but de ses recherches et accède au résultat en un clic.

b) Rendre cohérente les informations d'une l'entreprise : les informations provenant d'une branche de l'entreprise peuvent être mise en corrélation avec celles d'une autre branche. Si deux unités de mesure portent le même nom, elles doivent alors signifier la même chose. A l'inverse, deux unités ne signifiant pas la même chose doivent être définie différemment. Une information cohérente suppose une information de grande qualité. Cela veut dire que l'information est prise en compte et qu'elle est complète. Le Data Warehouse joue un rôle de référentiel pour l'entreprise puisqu' 'il permet de fédérer des données souvent éparpillées dans différentes bases de données

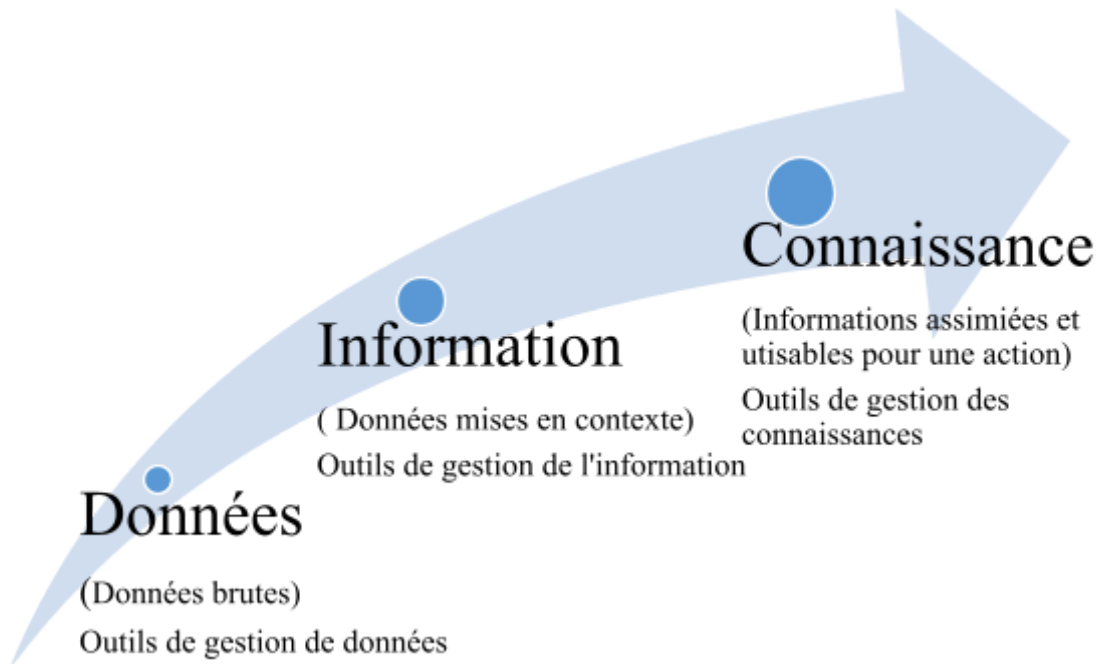
c) Constituer une source d'information souple et adaptable : l'entrepôt de données est conçu dans la perspective de notifications perpétuelle, l'arrivé de question nouvelles ne doit bouleverser ni les

données existantes ni les technologies. La conception de Data Mart distincts composant un entrepôt de données doit être répartie et incrémentielle.

d) Représenter un bastion sécurisé qui protège la capitale information : l'entrepôt de données ne contrôle pas seulement l'accès aux données, mais il offre à ses gestionnaires une bonne visibilité des utilisations.

e) Constituer la base décisionnelle de l'entreprise : l'entrepôt de données recèle en son sein les informations propres à faciliter la prise de décisions. Elle permet le développement d'applications décisionnelles et de pilotage de l'entreprise et de ses processus.

4.KNOWLEDGE MANAGEMENT



- La gestion des connaissances (en anglais Knowledge Management) est l'ensemble des initiatives, des méthodes et des techniques permettant de percevoir, d'identifier, d'analyser, d'organiser, de mémoriser, et de partager des connaissances entre les membres des organisations, en particulier les savoirs créés par l'entreprise elle-même (ex : marketing, recherche et développement) ou acquis de l'extérieur (ex : intelligence économique) en vue d'atteindre l'objectif fixé.
- Les acteurs d'une organisation ne doivent pas se limiter à la consommation d'informations brutes. Après avoir vérifié les informations (sources, origines), ils doivent veiller aux usages de celles-ci, ce qui signifie interprétation, structuration, capitalisation, et partage des connaissances.
- Dans ce cadre, les systèmes d'information constituent un élément central pour tout type de stratégie de gestion des connaissances.

3) Les objectifs de la gestion des connaissances

- ❑ Optimiser le partage des connaissances et favoriser les échanges entre les différents acteurs,
- ❑ Soutenir l'apprentissage individuel, collectif et organisationnel,
- ❑ Effacer le risque de perte des savoirs individuels,
- ❑ Restreindre le phénomène de Turnover.
- ❑ Optimisation de la production
- ❑ Aide à la décision
- ❑ Valorisation des compétences des individus

Augmentation de l'efficience organisationnelle

Inconvénients de la gestion des connaissances :

- ❑ Investissement coûteux
- ❑ Difficulté d'implantation
- ❑ Restreindre l'innovation
- ❑ les échanges de connaissances tacites sont conditionnés par la hiérarchie en place

3 types de connaissances

- ✓ Connaissance interne structurée (savoir)
- ✓ Connaissance extérieure
- ✓ Connaissance interne informelle (tacite, savoir-faire)

5.Business intelligence & DATAMINING

1- ENVIRONNEMENT DE BUSINESS INTELLIGENCE

1.1- Catégorie des données :

Les données rassemblées dans un entrepôt de données sont issues des applications transactionnelles (systèmes opérationnel), pour la plupart et de bases de données externes après avoir été extraite par ETL.

- Données de détails : issues directement d'un système transactionnel de l'entreprise. Malgré la capacité de stockage, on ne stocke que les données susceptibles d'avoir un sens dans le processus d'aide à la décision.
- Les données agrégées : sont construite à partir de données de détails afin de faciliter la navigation suivant les besoins décisionnels.
- Les métadonnées : (données des données) décrivent les caractéristiques des données stockées (origine, date ...). Elles sont utiles aussi bien pour l'utilisateur que l'administrateur

1.2- LES OUTILS DE BI :

Le processus BI vise à récupérer des données brutes (ERP, CRM ...), et à les transformer en information et à les diffuser sous forme de tableaux de bord ou reporting

Datawarehouse : c'est une base de données utilisé pour collecter et stocker les informations volatiles provenant d'autre base de données.

Datamart : sous ensemble logique d'un datawarehouse. Il est généralement exploité en entreprise pour restitués des informations ciblés sur un métier spécifique

ETL : il s'agit d'une technologie informatique permettant d'effectuer des synchronisations massives d'informations d'une base de données vers une autre.

Cube OLAP : représentation abstraite d'informations multidimensionnelles. Les données sont rangées selon un principe de dimensions correspondant étroitement aux axes de recherche des utilisateurs

1.3- ORGANISATION DE BI EN SEIN DE L'ENTREPRISE :

- Dans un contexte centralisé : il existe une stratégie BI, pour tout le groupe. Dans ce contexte l'activité BI vise à consolider les données prévenantes directement des différentes entités du groupe. Les entités ne disposent pas de d'outils BI, elles dépendent complètement de BI groupe
- Dans un contexte décentralisé : dans ce contexte, il existe deux BI.

Un BI de groupe_piloté par le groupe : pilotage des processus transversaux.

Un BI pour les entités opérationnelles, qui sert les besoins de ces entités.

1.4- ARCHITECTURE BI :

Une architecture de BI est un ensemble de concepts, outils, méthodes, et technologies (logicielles et matérielles) qui, une fois mises en relation, permettent de créer de la connaissance et répondre aux besoins stratégiques de l'entreprise. Une architecture de BI peut être comparée à la structure d'une maison, c'est l'ensemble des poutres, murs, tuyauterie, etc... Il est important de comprendre ce concept d'architecture et de savoir le différencier des autres termes tels qu'environnement, système, logiciel, etc.

2- FONCTIONNEMENT OPÉRATIONNEL :

La chaîne décisionnelle comprend différentes phases :

- La **phase d'alimentation** : Cette phase va faire intervenir des processus ETL qui se chargeront de récupérer toutes les données nécessaires depuis les différentes sources de stockage.
- La **phase de modélisation** : Cette phase est le moment où les données sont stockées sous une forme adaptée pour les analyses que nous souhaitons effectuer. Elle contient notamment

le datawarehouse chargé de centraliser les données. Elle fait aussi intervenir les notions de cubes et de datamarts nécessaire pour coller au plus près des attentes métiers.

- La **phase de restitution** : C'est dans cette phase que les différents outils de restitution vont intervenir. Nous trouverons des outils de reporting, des portails d'accès à des tableaux de bord, des outils de navigation dans des cubes, ou des outils de statistiques.
- La **phase d'analyse** : C'est dans cette phase que les utilisateurs finaux interviennent et analysent les informations qui leur sont fournies. Elle peut aussi faire intervenir des spécialistes en analyse pour utiliser des outils de statistique et ressortir des prévisions ou des estimations futures (datamining).

3- LES PROCESSUS ETL

Les processus ETL (Extraction, Transformation et Chargement) sont en charge de récupérer les données depuis l'ensemble des sources opérationnelles existantes et de les charger vers le système décisionnel :

- Extraction de données des bases de données opérationnelles (ERP, SGBDR, fichiers en dur, etc.)
- Transformation de ces données pour nettoyer, standardiser, documenter, corriger et de dupliquer.
- Chargement des données dans le système décisionnelles : Datawarehouse, Datamarts, ou Cube.

Pourquoi est-il intéressant d'utiliser un outil ETL ?

- Les données à traiter pour les systèmes décisionnelles sont généralement très volumineuses, les ETL sont adaptés à cette volumétrie en croissance exponentielle.
- D'autres part, les ETL proposent un grand nombre de composant permettant de se connecter facilement à tout type de source. Ils implémentent aussi des outils qui offrent des services annexes comme par exemple un envoi d'alertes pour prévenir les administrateurs en cas de problèmes.
- La représentation graphique des scripts offre une prise en main facile et rapide. Ces outils permettent de clarifier et standardiser les processus ce qui est un gain en matière d'évolution et de maintenance du système

I- Data mining

1- EVOLUTION DES BASES DE DONNÉES :

En 1960, y a eu la création des bases de données grâce à la collecte des données.

Dans les années 70, l'évolution des technologies de l'information a abouti au développement de techniques d'indexation, d'organisation et de manipulation des données. A la même époque, on a

inventé le modèle relationnel, qui devint vite un standard pour les bases de données. Peu à peu les interfaces s'améliorèrent et l'accès aux données fut facilité pour l'utilisateur grâce à l'introduction de langages de requêtes (SQL, System Query Language).

Puis, avec l'apparition de l'OLTP (On-Line Transaction Processing), un type d'opération informatique où l'ordinateur répond immédiatement à la requête, les bases de données relationnelles ont été rapidement reconnues comme un outil majeur de stockage, de classement et de restitution des données.

Dans les années 1990, et à l'issue de la technologie des data warehouse, y eut l'émergence du data mining, il était au départ destiné à aider à la découverte d'informations cachées (ou passées inaperçues) dans les bases de données.

Autre avantage par rapport à l'OLAP, ce qui différencie les deux outils est surtout la capacité du data mining à incorporer des techniques de compréhension intelligente des données. Contrairement à l'OLAP qui doit être guidé étape par étape et qui est entièrement contrôlé par l'utilisateur.

2- Data-Mining : les raisons du développement

Le développement du data mining est due à plusieurs raisons initialement :

- ☐ L'explosion de données : les outils automatiques pour la collecte des données ont fait que les bases de données ont augmenté en taille, ce qui a permis à l'entreprise de disposer d'une quantité énorme de données mais qui n'était pas exploitée d'une manière optimale.
- ☐ Développement du datawarehouse : Avec le développement "Data Warehouse" et des solutions de stockage, la quantité des informations détenues par les entreprises devient de plus en plus importante. Beaucoup d'informations pourtant particulièrement intéressantes peuvent ne pas apparaître à première vue.
- ☐ Développement des technologies de l'information : les entreprises se permettent de plus en plus de stocker toutes les informations grâce au faible coût de stockage de données, et aux saisies automatiques des transactions
- ☐ Amélioration de la productivité L'accroissement de la concurrence, l'individualisation des consommateurs et la brièveté du cycle de vie des produits obligent les entreprises à non simplement réagir au marché mais à l'anticiper.
- ☐ Prise des décisions stratégiques efficaces : Besoin accru d'avoir une vision globale et synthétique pour prendre des décisions stratégiques efficaces en temps réel aussi il faut exploiter le vécu (données historiques) pour prédire le futur et anticiper le marché.

Dans cette optique, l'utilisation de l'outil data mining offre plusieurs avantages aux entreprises, et surtout faire face à la multiplication de bases de données très importantes, souvent inutilisables par manque de temps ou de moyens.

3- CE QU'EST LE DATA MINING

Le *Data Mining* est en fait 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 Data Warehouse ou DataMart.

Les techniques mises en action lors de l'utilisation de cet instrument d'analyse et de prospection sont particulièrement efficaces pour extraire des informations intéressantes (non triviales, implicites, préalablement inconnues et potentiellement utiles) à partir de grandes bases de données.

L'outil de prospection Data Mining est à même de trouver des structures originales et des corrélations informelles entre les données.

Il permet de mieux comprendre les liens entre des phénomènes en apparence distincts et d'anticiper des tendances encore peu discernables.

4- COMMENT ON L'UTILISE ?

A contrario des méthodes classiques d'analyses statistiques, Cet instrument d'analyse est particulièrement adapté au traitement de grands volumes de données. Avec l'augmentation de la capacité de stockage des supports informatiques, un maximum de renseignements sera capté, ordonnés et rangés au sein du Data Warehouse.

Autres appellations: ECD (Extraction de Connaissances à partir de Données)/KDD (Knowledge Discovery from Databases), fouille de données, exploration des données

5- TYPES DES DATA MINING :

A- Classification selon les types de bases de données exploitées:

- Relationnelle
- Transactionnelle
- Orienté-objet
- Object-relationnelle
- Active
- Spatiale
- Séries temporelles
- Texte

- Multi-media
- Hétérogènes
- WWW

B- Classification selon les connaissances recherchées :

- Association
- Classification
- Clustering
- Tendances
- Analyse de déviation.

Multiples fonctions aux différents niveaux

C- Classification par domaine d'application :

Le Data mining a aujourd'hui une grande importance économique du fait qu'il permet d'optimiser la gestion des ressources (humaines et matérielles). Il est utilisé par exemple :

- Organisme de crédit : gestion et analyse des risques par exemple pour décider d'accorder ou non un crédit en fonction du profil du demandeur de crédit, de sa demande, et des expériences passées de prêts
- Optimisation du nombre de places dans les avions, hôtels, ... $\Rightarrow$ sur réservation
- Organisation des rayonnages dans les supermarchés en regroupant les produits qui sont généralement achetés ensemble (pour que les clients n'oublient pas bêtement d'acheter un produit parce qu'il est situé à l'autre bout du magasin). Par exemple, on extraira une règle du genre : les clients qui achètent le produit X en fin de semaine, pendant l'été, achètent généralement également le produit Y
- Médecine et pharmacie : f Diagnostic : découvrir d'après les symptômes du patient sa maladie pour choisir le médicament le plus approprié pour guérir une maladie donnée
- Commerce électronique, recommandation de produits
- Analyser les pratiques et stratégies commerciales et leurs impacts sur les ventes
- Moteur de recherche sur internet : Web mining: par exemple : Les logs des accès Web sont analysés pour découvrir les préférences des utilisateurs pour adapter l'interface à l'utilisateur/service.
- Extraction d'information depuis des textes : souvent désignée sous l'anglicisme text mining. C'est un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité, dans des textes produits par des humains pour des humains. Dans la pratique, cela revient à mettre en algorithmes un modèle simplifié des théories linguistiques dans des systèmes informatiques d'apprentissage et de statistiques

- Evolution dans le temps de données : fouille de séquences.
- Détection de fraudes : Utiliser les données historiques pour construire des modèles de comportements frauduleux puis utiliser les techniques de datamining pour retrouver des instances similaires

Exemple de la banque : Vous êtes à l'étranger et quelqu'un a volé votre carte de crédit ou votre mobile.

- la compagnie bancaire va utiliser les données historiques pour construire un modèle de comportement frauduleux et utiliser le data mining pour identifier des instances similaires.

D- Classification selon les techniques utilisées :

- BD
- Data warehouse (OLAP)
- Apprentissage
- Statistiques
- Visualisation

6.EDI

1. Définition de l'EDI

L'abréviation EDI signifie Electronic Data Inter change ou, en français, Echange de Données Informatisées.

L'EDI est le terme générique définissant un échange d'informations automatique entre deux entités à l'aide de messages standardisés. Dans la pratique, l'EDI permet de réduire significativement l'intervention humaine dans le traitement de l'information, ce dernier devient donc plus rapide et plus fiable. La rapidité et la meilleure fiabilité des échanges assurés par l'EDI permettent de fluidifier les informations échangées, ainsi tout en améliorant la sécurisation des transactions elles permettent de réduire considérablement les coûts de traitement. Ceci devient maintenant indispensable selon la politique de « gestion en juste à temps » qui augmente le nombre de transactions à traiter dans un temps de plus en plus court.

En bref, l'EDI est un échange direct standardisé, d'ordinateur à ordinateur, de documents d'affaires (ordres d'achats, mandats, paiements, analyses de stock, etc.) entre un émetteur, des fournisseurs et des clients ou partenaires.

Traitement d'informations sans EDI

Traitement d'informations avec EDI

2. Objectifs de l'EDI

L'idée de base derrière la conception de l'EDI était l'optique « zéro papier », qui implique une réduction notable des coûts (impression, courriers, fax, mise sous pli...)

Plus encore, grâce à la suppression des ressaisies, l'EDI permet de maximiser l'efficacité, de minimiser l'erreur, et donc permet de garantir une meilleure fiabilisation des données.

Par ailleurs, l'automatisation des processus a permis de réduire considérablement l'intervention humaine à ce niveau, une avancée qui a assuré une productivité plus importante.

Enfin, l'EDI reste un outil essentiel vers un meilleur suivi des échanges (commande, bon de livraison, facture...)

L'Exchange Data Inter change a pour rôle de satisfaire au mieux l'organisation de l'entreprise et permet l'échange des données de manière plus pratique et sécurisée tout en préservant les données. Bine entendu, c'est donc évidemment la raison pour laquelle on peut la caractériser d'indispensable pour ce qui est de la bonne marche de nombreuses organisations dans de différents secteurs

L'EDI est utilisé dans de nombreux cas :

- L'EDI s'avère être indispensable sur le niveau universitaire surtout pour ce qui est de l'envoi des différents documents aux étudiants (Emplois, relevé de notes,...),
- Pour ce qui est des fabricants, l'échange des données informatisées joue son rôle au niveau de la transmission des dessins techniques complexes,
- Les particuliers et les consommateurs utilisent de même un système EDI Lors de l'utilisation de la carte de crédit ou encore lors de leurs transactions avec leurs banques.
- les gouvernements commencent à y faire appel pour les déclarations d'impôt qui sont transmises par voie électronique.

L'EDI concerne de nombreux secteurs d'activité :

Il est tout de même nécessaire de dire que l'EDI est fortement répandu et fait partie désormais de nombreux domaines citons :

Les échanges commerciaux : entre les clients et les distributeurs.

Secteur des transports : le suivi du transport ainsi que la logistique.

La sous-traitance : Afin que l'entreprise puisse échanger les données avec son sous-traitant

la banque et la finance : entre les banques, entre les banques et la banque centrale,

La santé : produits pharmaceutiques

D'autres domaines : l'automobile, la distribution des aliments, la banque et la finance, le bâtiment, la justice, la grande distribution, le secteur social, le tourisme..

III. Les bénéfices et les freins liés à l'utilisation

Bénéfices

L'EDI a pour mission d'échanger de multiples données structurées à partir d'un ordinateur vers un autre via des messages précédemment préétablis et normalisés, cela se fait à travers un mode de communication électronique, entre partenaires indépendants.

L'EDI est considéré comme étant un avantage concurrentiel pour l'entreprise, c'est un atout permettant une meilleure organisation. Ce dernier facilite le travail vu qu'il remplace toute opération manuelle (commandes, factures, avis d'expéditions, etc..). Cela dit, grâce à l'EDI, l'entreprise peut échanger de multiples informations structurées sans pour autant que l'objet soit des documents sous une forme

papier. Ce type d'informations est ainsi nommé « message », ce type d'échange rend la tâche moins compliqué et forcément plus pratique.

Exemple : une facture électronique échangée en EDI entre directement dans l'outil de gestion, elle est donc « intégrée », alors qu'une facture papier fera l'objet d'une saisie dans le système de gestion.

L'EDI a de nombreux avantages :

- Accélération du cycle de la transaction commerciale et des règlements
- Réduction des coûts administratifs
- Fidélisation du partenaire commercial
- Réduction de la distance entre distributeur et client
- L'EDI permet à l'entreprise d'élargir le choix et l'offre de ses produits
- Meilleure correspondance entre flux physiques et flux d'informations
- Prestations logistique et transport
- Améliorer la gestion des stocks et la gestion de trésorerie
- Commodité de l'accès à l'information
- Améliorer la structuration des archives
- Fiabilité et véracité des statistiques
- Meilleure gestion temporelle et spatiale

D'autres avantages Provenant de l'utilisation de la norme EDIFACT/ONU :

- Interfaçage unique avec tous les correspondants (partenaires commerciaux, banques, transporteurs, . .)
- Interfaçage automatique avec les applications de l'entreprise.
- Saisie unique pour tout le cycle de la transaction
- Indépendance vis-à-vis des réseaux de télécommunications
- Facilité de communication avec les partenaires même à l'étranger (Mise en œuvre plus rapide grâce au standard international)

Freins liés à l'utilisation

En premier lieu, d'importantes connaissances techniques et informatiques sont exigées pour l'installation d'un logiciel EDI. Ensuite, vient intervenir l'inconvénient relatif aux dépenses nécessaires à la formation du personnel aptes à la manipulation d'un tel logiciel. Le système EDI nécessite une supervision à plein temps. D'autres parts, cette méthode est utilisée davantage lorsqu'il s'agit d'une opération avec un nombre de commandes élevés

Le coût reste assez élevé pour les PME : Les PME n'ont pas forcément un trafic de données informatisées suffisant pour justifier des dépenses en EDI. Cependant, avec Internet, on peut espérer une simplification de l'EDI. Toutefois, si Internet constitue une interface informatique intéressante pour les donneurs d'ordres, les petites entreprises seront en fait souvent amenées à réaliser la saisie des données informatiques manuellement, directement sur le site du client. Ainsi, seules les grosses structures éviteront des temps de saisie superflus.

7.WORKFLOW

1-1. Qu'est-ce qu'un WORKFLOW ?

En effet, il s'agit d'un anglicisme pour « flux des travaux », ou un processus permettant d'automatiser les flux d'information dans une entreprise. Plus profondément le WORKFLOW permet la modélisation des processus dans le cadre d'une démarche plus globale BPM Business Process

Management. Ainsi, on qualifie de Business Process (processus métier) l'ensemble des activités et de procédures qui permettent collectivement la réalisation d'un objectif métier. Le WORKFLOW correspond à l'automatisation de ce Business Process

Divers acteurs sont touchés par le processus et notamment les salariés mais aussi les personnes en charge de tâches extérieures à l'entreprise mais concernées par elles.

Les outils informatiques de WORKFLOW permettent de définir le parcours de chaque type de document, de contrôler les délais et d'en suivre précisément l'exécution. L'outil de WORKFLOW est suffisamment souple pour permettre de modifier au cours du parcours le circuit prédéfini afin d'éviter toute forme de blocage.

1-2. Les types des WORKFLOWS:

On peut aussi regrouper les WORKFLOWS selon les catégories suivantes :

- le Workflow de production : qui correspond à la gestion des processus de base de l'entreprise. Les procédures supportent peu de changements dans le temps, et les transactions sont répétitives (ex : contrats d'assurance, gestion de litiges, de réclamations clients, etc.),
- le Workflow Administratif : qui correspond à tout ce qui est routage de formulaires,
- le Workflow Ad-Hoc: pour la gestion des procédures non déterminées, ou mouvantes,
- le Workflow Coopératif : gérant des procédures évoluant assez fréquemment, et liées à un groupe de travail restreint dans l'entreprise.

1-3. Les rôles de l'outil Workflow

- 🕒 Il partage à chaque intervenant l'information nécessaire à l'exécution de ses tâches.
- 🕒 Il signale les délais de livraison ou les délais d'accomplissement de la tâche
- 🕒 et rappelle les modes de validation.
- 🕒 Ses différentes fonctions lui permettent d'organiser les différentes tâches que les salariés doivent accomplir et de s'assurer que chaque tâche est bien réalisée.
- 🕒 Il contribue à l'amélioration de la productivité.
- 🕒 Il réduit la communication écrite.
- 🕒 Il renforce le volume de sécurité et de confidentialité car le WORKFLOW supervise et contrôle les droits et fonctions de chacun.
- 🕒 Un outil de rationalisation, notamment à travers la redéfinition des processus de gestion de l'entreprise afin de supprimer les redondances, standardiser les méthodes de travail, s'adapter rapidement aux évolutions.

Ce qui permet au manager de mieux contrôler la bonne marche, la performance et déterminer les points de blocage.

3-1 Les apports du workflow

Les apports du workflow sont très nombreux. Les plus importants sont les suivants :

- 🕒 Gain de place : grâce à la diminution des archives papier.
- 🕒 Gain de temps et de réactivité : recherche instantanée des documents.
- 🕒 Gain de qualité : pas de doublons, les documents sont à jour ...
- 🕒 Gain de mobilité : on accède aux documents depuis n'importe quel poste.

🌐 Gain de sécurité et de confidentialité : droits d'accès des utilisateurs définis en fonction du type de document.

🌐 Automation : La répétition continue de ces procédés, bien qu'ils soient simples (pensez par exemple à la saisie de données) peut prendre un temps incroyable. En outre, une telle répétition est impossible lorsque vous vous retrouvez à employer des talents créatifs et variés dans votre équipe. Avec un outil de gestion de workflow professionnel, ces processus peuvent être automatisés et complétés en quelques secondes seulement.

🌐 Collaboration : La collaboration avec des parties externes peut entraver ou retarder le processus de production et la mise sur le marché de vos projets sous forme de fichiers qui doivent constamment être remplacés et approuvés par les différents actionnaires. Avec un outil de Workflow Management, cependant, ces étapes peuvent être intégrées dans un processus standard, effectuées simultanément, et gérées dans un outil centralisé.

🌐 Communication : Afin de s'assurer que toutes les instructions et commentaires soient vus et revus, des notes et annotations peuvent être nécessaires. Un bon logiciel de Workflow Management vous fournira ces outils, et vous pouvez être sûr que tous les participants seront à jour sur tous les développements.

🌐 Distribution : Envoyer de grands fichiers numériques à des collègues est à la fois long et stressant, peu importe la méthode. La plupart des boîtes de réception de courriel refusent de recevoir des pièces jointes de 25 Mo ou plus, ce qui est un montant forfaitaire si vous envoyez des images de fichiers numériques bruts (une nécessité quotidienne pour tous les gestionnaires de marque ou marketeurs). Un outil de gestion de workflow vous permet de partager des fichiers volumineux directement avec les autres, sans avoir recours à des méthodes pénibles et précaires, comme l'e-mail ou le FTP.

🌐 Duplication : Il est commun de retravailler un fichier à plusieurs reprises afin d'économiser sur les ressources; en réalité, cela mène au chaos pour savoir quelle est la dernière version. En conséquence, les anciennes versions se font utiliser, et les nouvelles versions se perdent dans la masse. L'efficacité de la gestion des workflows est garantie - grâce à une gestion de version intelligente, personne ne travaille avec une version obsolète.

🌐 Vue d'ensemble : Lorsqu'on travaille sur différents projets à la fois, il peut être difficile de garder une trace de ce qui se passe et où cela se passe. Avec un outil de gestion de workflow, l'information est pertinente et est clairement affichée sur le tableau de bord, si toutes les tâches et les statuts du projet sont clairement visibles

🌐 Rationalisation: suppression des redondances, standardisation des méthodes de travail, adaptation rapide aux évolutions

Sécurisation : réduire les risques d'erreurs humaines, améliorer la traçabilité et la transparence de l'information

🌐 Information en temps réel : visibilité immédiate des tâches effectuées, en cours, à faire ou en retard, anticipation des blocages et dysfonctionnement

8.SUPPLY CHAIN MANAGEMENT

A. Définition Logistique

Le terme de logistique appartient au monde militaire et désigne ainsi la gestion au mieux du ravitaillement, du logement et du transport des troupes. Par la suite le terme a été étendu au monde de l'entreprise pour désigner la gestion des flux matières et matériel, en entrée comme en sortie, des stocks et du transport des produits.

B. Définition Supply Chain

Supply chain est le terme anglo-saxon utilisé pour désigner la chaîne d'approvisionnement. Ce terme correspond peu ou prou à celui de logistique vu ci-dessus. Dans la pratique, lorsque l'on choisit d'utiliser le terme de supply-chain en lieu et place du vocable traditionnel en France de logistique, c'est bien souvent dans l'esprit du Supply Chain Management, une notion bien plus complexe.

C. Définition Supply Chain Management

Le Supply Chain Management ne désigne pas uniquement une famille de produits logiciels d'entreprise destinés à faciliter la gestion de ladite chaîne, connus sous le sigle SCM, il existe quatre conceptions différentes du Supply Chain Management.

🌐 Pour les « Traditionnalist », le SCM est une fonction ou un sous-ensemble de la logistique, plus particulièrement en lien avec les partenaires de l'entreprise (fournisseurs et clients).

🌐 Pour les « Re-labeling », le SCM n'est rien d'autre que la logistique. Il y a simplement un changement de nom dans une démarche très marketing de la fonction. Certains s'inscrivent également dans ce groupe mais en considérant, qu'actuellement, les pratiques Supply Chain des entreprises ne sont rien d'autres que de la logistique. C'est, dans ce cas, une adhésion plus par défaut que par conviction.

🌐 Pour les « Intersectionnist », le SCM fait le lien entre les fonctions achats, logistique, marketing... et apparaît, dans les entreprises, sous la forme d'une équipe qui conseille plus qu'elle n'agit directement sur les flux (vision très fonctionnelle).

🌐 Pour les « Unionist », la logistique est une fonction du SCM au même titre que les achats, la gestion industrielle ou le marketing. Il doit exister un département Supply Chain qui gère aussi bien sur le plan opérationnel que stratégique le pilotage des flux de l'entreprise.

La plupart des réflexions des Universitaires et Chercheurs adhèrent totalement à la perspective « Unionist » qui considère donc la logistique comme un sous-ensemble de la Supply Chain. Ce rattachement théorique nous conduit donc à adopter la définition suivante :

Le SCM comprend la prévision et le management de toutes les activités relevant de la recherche de fournisseurs, de l'approvisionnement, de la transformation et toutes les activités du management logistique.

De façon essentielle, il inclut la coordination et la coopération avec les partenaires de la chaîne qui peuvent être les fournisseurs, les intermédiaires, les prestataires de services et les clients. Par essence, le SCM intègre le management de l'offre et la demande dans et entre les entreprises.

Comme il faut différencier entre le SCM et le Management Logistique qui est qu'une composante du management de la Supply Chain qui prévoit, met en place et maîtrise de façon efficiente et efficace

les flux aller et retour de marchandises, leur entreposage et des services grâce à des informations associées, de manière à satisfaire les exigences du client.

Il s'agit là alors de planification et d'exécution logistique. Nous sommes aujourd'hui convaincus que parmi nos adhérents, différentes cultures d'entreprises et positionnements de la Supply Chain existent mais aussi évoluent.

2) Les facteurs d'efficience et d'efficacité du Supply Chain Management

Le supply chain management qui se résume la gestion des flux physiques et d'information du client du client au fournisseur du fournisseur, afin d'offrir une réponse la plus satisfaisante possible aux besoins des clients. Les raisons d'émergence, d'efficience et d'efficacité de la gestion de la chaîne logistique se subdivisent à des facteurs internes et d'autres qui sont externes.

a) Les facteurs internes d'efficience et d'efficacité du SCM:

- a. La baisse des coûts de transaction avec l'amélioration des infrastructures de transport donnant l'accès à des zones économiques « à prix bas » (pays de l'Est, Chine, l'émergence des nouvelles technologies de communication,
- b. Le recentrage sur de l'activité sur un maillon de la chaîne de la valeur où elle permet de développer une compétence distinctive et un avantage concurrentiel,
- c. La diminution des coûts de gestion de stocks par la diminution dans toute la chaîne logistique,
- d. La création de valeur partenariale avec un réseau performant de fournisseurs et de sous-traitants,

b) Les facteurs externes d'efficience et d'efficacité du SCM:

- a. L'amélioration de la qualité globale des produits qui s'est accrue en utilisant des méthodes de la qualité totale etc..
- b. La coopération « éthique » entre les acteurs, la morale des affaires, l'éthique est indispensable au bon fonctionnement de la supply chain et de l'entreprise en réseau. L'entreprise étendue ne peut exister qu'à partir d'un certain niveau de confiance indispensable.
- c. Le rôle fondamental des systèmes d'information et des technologies de l'information et de la communication (TIC), les réseaux et la communication, le développement de standards de communication, d'une nouvelle forme d'infrastructure (Web, EDI,...) ont offert à la flexibilité quatre propriétés essentielles la compression du temps, l'espace. Le web permet de dépasser la relation de type B2B, imposée par l'EDI, et d'aller vers une relation B to C voire C to C. C'est une nouvelle vague d'innovation technologique qui vient après celle de la puissance de calcul et de la micro-informatique des années 80.

3) La structure du SCM:

La structure du SCM comprend la combinaison de trois éléments inter-reliés :

☛ La structure relationnelle de la supply chain (supply chain network structure) qui permet d'identifier les membres (fournisseurs, clients, prestataires), les différents niveaux (rangs) et la stratégie de gestion et d'intégration ;

☛ Les processus managériaux de la supply chain (supply chain business processes) constitués par des milliers d'activités et qui requièrent une intégration. Par exemple : la gestion de la relation clientèle, la gestion de la production...

☛ Les composantes de management de la supply chain (supply chain management components) qui regroupent les composantes physiques et techniques, qui gèrent et suivent les flux de produits et d'informations associés et les composantes managériales et comportementales, qui portent sur les facteurs intangibles tels que le pouvoir ou le leadership.

4) Les outils du SCM:

Il s'agit d'un processus où chaque étape est gérée par ses propres outils.

a. La première étape étant la planification peut être gérée par MRP, JIT, DRP.

MRP : (Material Requirements Planning) est une méthode de gestion et de planification de production en flux poussé basé sur les nomenclatures de produits et les prévisions des ventes.

Le principe du MRP consiste à déterminer de manière automatisée, pour chaque produit fini, à partir du besoin brut, le besoin net en tenant compte des ordres prévisionnels et des ordres lancés ou fermes.

JIT : Le juste à temps est une méthode d'approvisionnement qui consiste à se faire livrer les matières ou produits au moment exact du besoin pour une utilisation directe. Son but principal repose essentiellement sur la volonté de :

- o Réduire les coûts logistiques d'approvisionnement, de production et de distribution,
- o Eviter toute forme de gaspillage,
- o Réduire les stocks de matières et les stocks de produits finis ,
- o Réduire les stocks de production.

DRP : (Planification des ressources de distribution) Méthode de calcul des quantités à approvisionner par référence de manière à éviter les ruptures tout en limitant les niveaux de stocks. Cette méthode permet de dimensionner les ressources logistiques et financières nécessaires.

b. La deuxième étape étant la fabrication repose aussi sur certains outils:

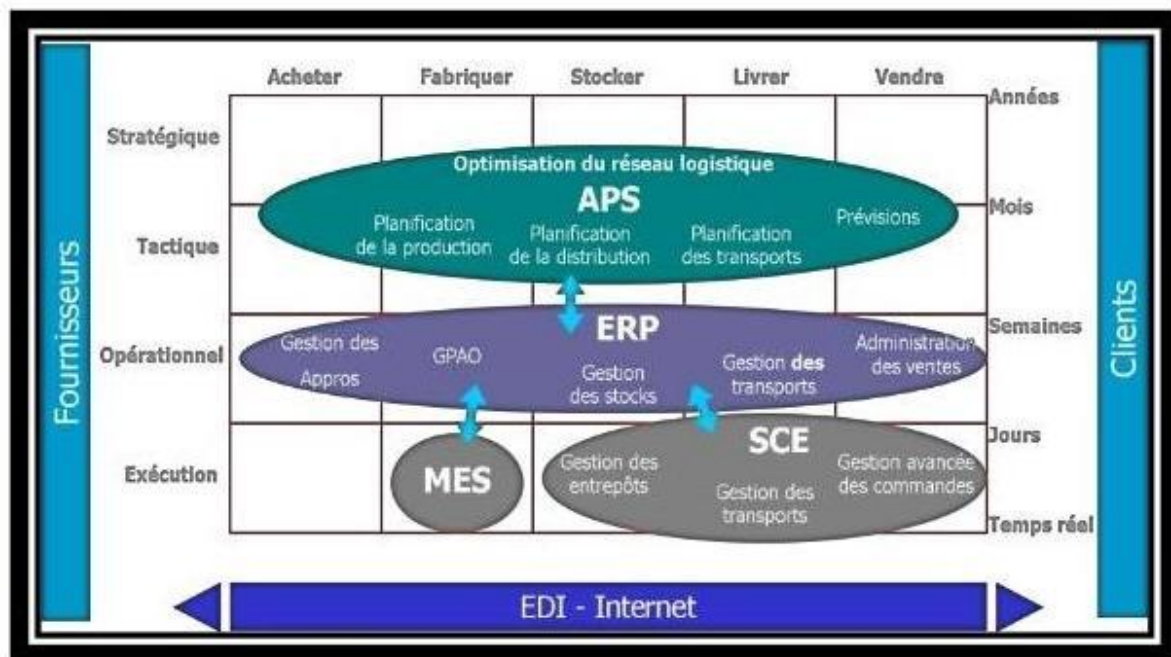
KANBAN : Cette méthode, déployée à la fin des années 1950 dans les usines Toyota, est mise en place entre deux postes de travail et limite la production du post amont aux besoins exacts du poste aval. Cette méthode est surtout adaptée aux entreprises ayant une production répétitive et relativement régulière.

OPT : La méthode OPT est une démarche de progrès permanente. Cette méthode est applicable à tous les processus de production caractérisés par des goulets d'étranglement (tâches critiques pour lesquelles un retard n'est pas rattrapable : un retard sur cette tâche retarde d'autant la fin d'un projet). Cette méthode répond à trois objectifs :

- ☛ maximiser les ventes,
- ☛ réduire les stocks;
- ☛ réduire les coûts.

CRP : Programme d'approvisionnement continu (Continuous replenishment program - CRP) : Programme qui déclenche la fabrication et le mouvement d'un produit dans la chaîne logistique d'approvisionnement dès qu'un produit identique est acheté par le consommateur final.

- c. La Troisième étape étant le stockage.
- d. La quatrième étape et la dernière est le transport.



Niveau stratégique : concerne les décisions de planification à long terme qui impliquent une prise de risque élevée de la part de l'entreprise :

- Localisations, missions et relations des usines et entrepôts :
- Positionnement des nouvelles usines et fermetures de sites,
- niveaux de capacité des usines et entrepôts,
- acquisitions de biens technologiques et d'équipements
- conception d'usines et d'entrepôts,
- décision des répartitions entre « faire et faire faire »
- réseaux de transports et prestataires de transport,

Niveau tactique : s'intéresse aux décisions à moyen terme qui devront être mises en application pour développer la stratégie décidée par l'entreprise. Elles tiennent compte des délais liés au cycle complet de production (englobant le délai d'approvisionnement de fabrication et de mise sur le marché).

- Affectation des capacités de production aux familles de produits par usine, par périodes temporelles de taille moyenne : trimestrielle,
- Taux d'utilisation des capacités planifiées par site et au niveau du réseau,
- Besoins en ressources (niveaux nominal et d'heures supplémentaires),
- Allocations des sources d'approvisionnement aux sites,
- Gestion des transferts intersites,
- Modes de transferts et choix des transporteurs.

Les APS : (Advanced Planning Systems) qui sont des progiciels de planification et d'optimisation qui se basent sur une prévision consolidée des ventes.

Niveau opérationnel : les décisions prises à ce niveau auront une portée limitée dans l'espace et dans le temps (court terme). Elles se situent au niveau d'un site voire d'un atelier : suivi et contrôle de la production, des commandes...

- 🌐 Ordonnancement quotidien et hebdomadaire et gestion des priorités,
- 🌐 Gestion des stocks à court terme,
- 🌐 Gestion et ordonnancement des commandes clients,
- 🌐 Ordonnancement et gestion des entrepôts,
- 🌐 Ordonnancement des tournées de véhicules.

Les ERP : sont des progiciels de gestion intégrée. Ils sont découpés en modules correspondants à des ensembles cohérents de fonctionnalités : module de gestion de la production, de l'approvisionnement et des stocks / module comptable, achats et force de vente / module de comptabilité générale, de comptabilité analytique...

Les SCE : (Supply Chain Execution) qui correspond à des progiciels d'optimisation de la logistique de distribution.

Les MSE : (Manufacturing Execution Systems) qui sont des solutions pour le contrôle et l'optimisation de l'exécution et de la production.

9.LA SOA :

L'architecture orientée services est une forme d'architecture de médiation qui est un modèle d'interaction applicative qui met en œuvre des services (composants logiciels)

- ☞ avec une forte cohérence interne (par l'utilisation d'un format d'échange pivot, le plus souvent XML) ;
- ☞ Des couplages externes « lâches »(par l'utilisation d'une couche d'interface interopérable, le plus souvent un service web).



L'architecture orientée services répond aux problématiques informatiques des entreprises en termes de :

- ☞ Flexibilité
- ☞ Réutilisabilité
- ☞ Interopérabilité
- ☞ Réduction du couplage entre les applications

Au niveau organisationnel

Au niveau Technique

- Flexibilité technologique
- Contre la complexité
- Flexibilité économique
- Possibilité d'ajout de nouveaux services
- La réutilisabilité des ressources
- La simplification des problèmes de maintenances

Réduction des coûts

Coûts de propriété

Les ressources réutilisables