



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ MATHÉMATIQUES ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : réseau et télécommunication

Par :

**Chouchaoui ahmed amine
Dahmour youcef**

Sur le thème

Passage à l'échelle dans le filtrage via le cloud computing

Soutenu publiquement le 08/07/2019 à Tiaret devant le jury composé de :

Mr OUARED Djilali	Université ibn Khaldoun	Président
Mr KHAROUBI Sahraoui	Université ibn Khaldoun	Encadreur
Mr KOUADRIA Abderahmen	Université ibn Khaldoun	Examineur

REMERCIEMENTS

Tout d'abord merci à dieu pour tous

Nous adressons nos remerciements les plus sincères à notre promoteur MR KHAROUBI SAHRAOUI, pour nous avoir proposé ce sujet et de nous avoir conseillé et dirigé tout au long de l'année.

Merci à tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste travail, pour leurs conseils, leurs encouragements et leurs soutiens.

Enfin, nous tenons à remercier le jury pour nous avoir honoré en acceptant de juger notre travail.

Résumé

Les systèmes de recommandations sont utilisés avec succès pour fournir des items pertinents (exemple : les films, la musique, les livres, les nouvelles, les images) adaptées aux préférences des utilisateurs. Parmi les approches proposées, nous focalisons sur l'approche de filtrage collaboratif qui consiste à trouver l'information qui satisfait l'utilisateur en utilisant les évaluations des autres utilisateurs. La problématique traitée dans notre thème est le problème du passage à l'échelle qui limite fortement la performance de la recommandation.

Mots clé : système de filtrage, filtrage collaboratif, cloud computing, passage à l'échelle.

Abstract

Recommendation systems are used successfully to provide relevant items (eg, movies, music, books, news, images) tailored to user preferences. Among the proposed approaches, we focus on the collaborative filtering approach of finding information that satisfies the user by using other users' ratings. The problematic dealt with in our theme is the scaling up problem which strongly limits the performance of the recommendation.

Keywords: filtering system, collaborative filtering, cloud computing, scalability.

ملخص

يتم استخدام أنظمة التوصية بنجاح لتوفير العناصر ذات الصلة (مثل الأفلام والموسيقى والكتب والأخبار والصور) المصممة لتفضيلات المستخدم. من بين الطرق المقترحة، نركز على طريقة التصفية التعاونية لإيجاد المعلومات التي ترضي المستخدم باستخدام تقييمات المستخدمين الآخرين. والمشكلة التي يتم تناولها في موضوعنا هي مشكلة التوسع التي تحد بشدة من أداء التوصية

كلمات الأساسية: نظام التصفية، التصفية التعاونية، الحوسبة السحابية، التدرجية.

Sommaire

LISTE DES FIGURE	7
Liste des tableaux	8
Introduction générale.....	9
Chapitre n ^o 1 : filtrage de l'information	10
1. Introduction	11
2. Définition du filtrage d'information	11
3. Concepts de base du filtrage d'information.....	13
4. Recherche d'information versus filtrage d'information.....	13
4.1 Fonctionnement de base	13
4.2 Différence entre la recherche et le filtrage.....	14
5. Types de filtrage d'information	16
5.1 Filtrage basé sur le contenu	16
5.1.1 Définition.....	16
5.1.2 Les limits	16
5.2 Filtrage collaboratif	17
5.2.1 Définition.....	17
5.2.2 Architecture générale	18
5.2.3 Calcul de la prédiction pour un système de filtrage collaboratif :	19
5.2.4 Systèmes de filtrage collaboratif.....	23
5.2.4.1 Evaluation des recommandations.....	24
5.2.4.2 Formation des communautés.....	24
5.2.4.3 Production des recommandations	24
5.2.4.4 Profils et communautés	24
5.2.5 Le filtrage collaboratif et le basé sur le continu	25
5.2.6 Limites du filtrage collaboratif	25
5.3 Filtrage Hybride.....	26

6. Conclusion.....	27
Chapitre n°2 : services du cloud computing.....	28
1. Introduction	29
2. Passage à l'échelle en filtrage d'information (Scalability)	29
3. Cloud computing.....	30
3.1. Définition	30
3.2. Les caractéristiques essentielles du Cloud computing.....	31
3.3. Centre de données (Datacenter)	33
3.4. Types de Cloud Computing.....	33
3.5. Niveaux de service du Cloud Computing	34
3.5.1. Le logiciel en tant que service (SaaS)	35
3.5.2. Plateforme en tant que service (PaaS).....	35
3.5.3. Infrastructure en tant que service (IaaS)	35
3.6. Avantages et inconvénients des services.....	37
3.7. Modèle d'application de Cloud Computing	38
3.7.1. Phase stratégique.....	38
3.7.2. Phase de planification	39
3.7.3. Phase de déploiement	39
3.8. Avantages du Cloud Computing	39
3.9. Migration des données.....	40
3.10. Fournisseur de services cloud.....	41
3.10.1. Amazon Web Services (AWS).....	41
3.10.2. Google Cloud.....	42
3.10.3. IBM Bluemix	42
3.10.4. Alibaba Cloud	43
3.10.5. Oracle Cloud	43
3.10.6. Microsoft Azure.....	43
3.11. Limites du Cloud Computing	44
4. Multithreading.....	45

4.1.	Définition	45
4.2.	Avantage du multithreading	45
4.3.	Inconvénients	45
4.4.	Multithreading versus Multi-processing	46
5.	Conclusion :	48
Chapitre n°3 : Expérimentation		49
1.	Environnement de travail	49
1.1.	Technologie utilisées	49
1.1.1.	Java	49
1.1.2.	JavaFx	49
1.1.3.	CSS	49
1.1.4.	FXML	50
1.2.	Présentation des outils de développement	50
1.2.1.	Netbeans	50
1.2.2.	SceneBuilder	51
1.3.	Bibliothèque utilisée	51
1.3.1.	JDK	51
1.3.2.	FontAwesome-Fx	51
1.3.3.	JFoenix	51
2.	jeu de données	51
3.	Mesures de tests	52
4.	Résultats	52
5.	Conclusion	59
Conclusion générale		49
Bibliographes		50
Webographie		51

LISTE DES FIGURE

Figure 1 : Schéma général de filtrage d'information.....	12
Figure 2 : Etape d'un système de recommandation.....	12
Figure 3 : modèle de recherche d'information	13
Figure 4 : modèle de filtrage d'information.....	14
Figure 5 : Principe général de filtrage collaboratif.....	17
Figure 6 : filtrage collaboratif.	18
Figure 7 : Architecture générale d'un système de filtrage collaboratif.....	19
Figure 8 : Les méthodes de filtrage collaboratif.	20
Figure 9 : Trois processus principaux d'un système de filtrage collaboratif.	23
Figure 10 : filtrage hybride.	26
Figure 11 : caractéristique du cloud computing	32
Figure 12 : Types de cloud computing	34
Figure 13 : services du cloud	34
Figure 14 : Niveau de service cloud computing.	36
Figure 15 : Modèle d'application de cloud computing.	38
Figure 16 : fournisseur de cloud	44
Figure 17 : multiprocessing en dual-core système	46
Figure 18 : multithreading en dual-core système.	47
Figure 19 : technologie utiliser	49
Figure 20 : outils de développement.....	50
Figure 21 : bibliothèque utilisée	51
Figure 22 : page d'accueil	53
Figure 23 : page chargement de données.....	53
Figure 24 : choix du fichier rating.csv	54
Figure 25 : matrice d'évaluation.....	54
Figure 26 : Choix de la taille de la matrice.....	55
Figure 27 : l'interface du résultat.....	56
Figure 28 : MAE	57
Figure 29 : comparaisons entre des défèrenttes configurations	57
Figure 30 : l'interface de recommandation.....	58
Figure 31 : Liste de recommandation	59

Liste des tableaux

Tableau 1 : Comparaison entre recherche d'information et filtrage d'information.	15
Tableau 2 : Avantages et inconvénients des services [9] [13].	37
Tableau 3 : la déférence entre multithreading et multiprocessing	47
Tableau 4 : Tableau de description des champs du fichier rating.csv	52
Tableau 5 : configuration materielles	52

Introduction générale

Être informé étant une nécessité professionnelle et citoyenne, recevoir des informations ayant un certain niveau d'intérêt individuel permet à chacun d'apprendre, d'analyser, de critiquer toute nouvelle source d'information. Ainsi recevant toute nouveauté, l'utilité du filtrage permet donc d'éviter de procéder régulièrement à une recherche d'éventuelles avancées, cela procure à l'utilisateur bien évidemment une économie d'effort mais également une certaine sérénité.

Dans ce mémoire nous abordons l'inconvénient du passage à l'échelle (scalability) dans les systèmes de filtrage d'information. Ce problème est très fréquent et étudié par une large communauté des chercheurs dans le domaine recommandation. Lors du passage à l'échelle du Web le nombre d'utilisateur s'élève à l'ordre de milliers voire millions et de même pour les items. Or, le calcul de voisinage dans le cas de filtrage collaboratif doit passer une grande quantité de données en exécution et prend un temps considérable pour extraire la valeur de prédiction.

Organisation du document

Ce mémoire est constitué de trois chapitres.

Le premier chapitre est consacré à la présentation des concepts de base du filtrage d'information, tout en expliquant la différence par rapport à la recherche d'information. Le deuxième chapitre est dédié à la présentation de la solution cloud computing ainsi que le concept du multithreading. En dernier chapitre nous focalisons sur la partie expérimentations et résultats pour démontrer notre étude.

Chapitre n°1 : filtrage de l'information

1. Introduction

Avec l'explosion de la masse d'information sur l'Internet, il est devenu plus que jamais indispensable d'organiser ce gros volume d'information presque illimité afin d'en tirer le meilleur profit. Les recherches actuelles visent à concevoir des mécanismes capables de faire parvenir continuellement à l'utilisateur l'information qui l'intéresse au lieu que celui-ci dernier dépense son temps à la chercher, c'est la fonction du filtrage d'information.[5]

Le filtrage est un processus naturel car l'être humain de tout temps et de part sa nature à tendance à ne pas prendre et accepter toujours tout ce qui lui est soumis, mais cherche plutôt ce qui convient et répond à son besoin.

Avec l'expansion de Web et l'avènement du document électronique le filtrage de l'information est devenu impossible avec les seuls efforts de l'être humain, vu ses limites dans la maîtrise de cette masse presque illimitée d'information et du grand nombre de document qui circule sur la toile. D'où l'idée de déléguer la fonction de filtrer automatiquement l'information par des systèmes misent place à cet effet.

2. Définition du filtrage d'information

Le filtrage d'information est le processus permettant à partir d'un large volume d'informations dynamiques, d'extraire et de présenter les seuls documents intéressants un utilisateur ou un groupe d'utilisateurs ayant des centres d'intérêts relativement semblables. Ces intérêts sont préalablement décrits en besoins d'informations, appelés profils.

Le filtrage de l'information est donc un processus qui fait parvenir continuellement les informations aux personnes qui en ont besoin à partir d'un grand volume d'information dynamique, au lieu de les laisser dépenser leur temps à chercher ce qui les intéressent.

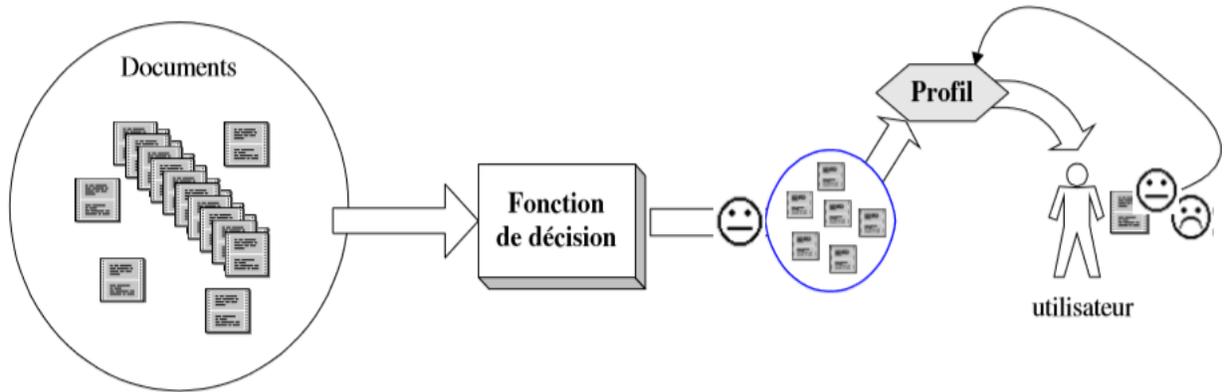


Figure 1 : Schéma général de filtrage d'information

Nous pouvons aussi définir le filtrage d'information comme une forme d'un système de recommandation visant à présenter les éléments d'information (films, musique, livres, news, images, pages Web, etc.) qui sont susceptibles d'intéresser l'utilisateur.

Un système de recommandation requiert généralement 3 étapes [3] :

Etape 1 : consiste à recueillir de l'information sur l'utilisateur.

Etape 2 : consiste à bâtir une matrice ou un modèle utilisateur contenant l'information recueillie.

Etape 3 : consiste à extraire à partir de cette matrice une liste de recommandations.

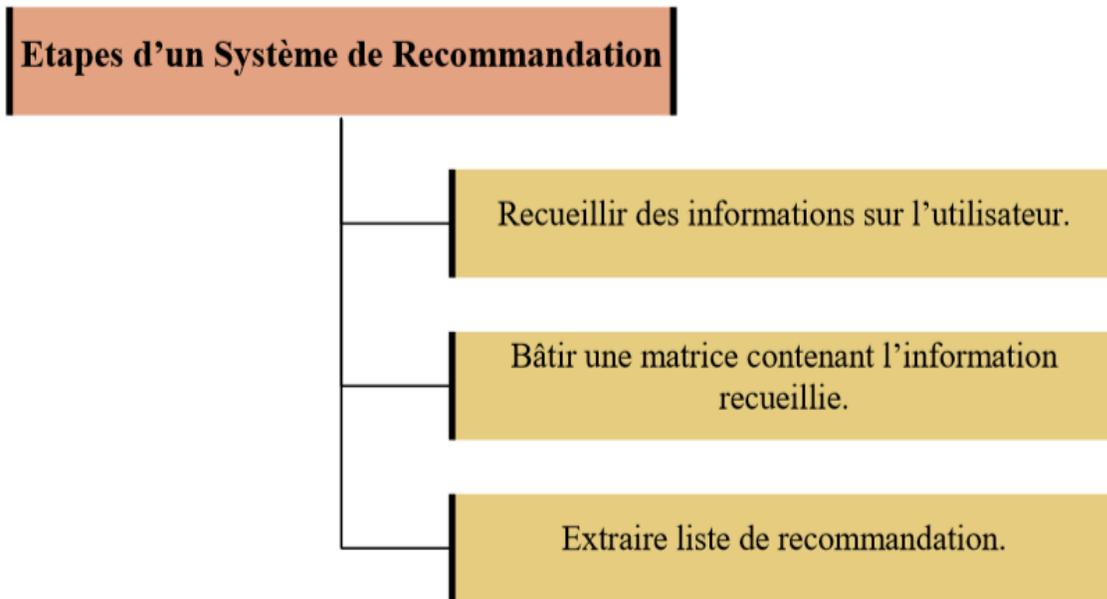


Figure 2 : Etape d'un système de recommandation.

3. Concepts de base du filtrage d'information

D'abord, Le filtrage d'information est concerné par des usages répétitifs du système, par une personne ou des personnes avec des buts et des intérêts à long terme donc il fait lui l'hypothèse que l'évolution des profils peut compenser ces problèmes, ensuite, les systèmes de filtrage envoient des documents à des groupes ou à des individus demeure la priorité fonctionnelle de plus il sélectionne ou élimine des documents à partir d'un flux dynamique de données.

Le filtrage est basé sur des descriptions des préférences d'un individu ou d'un groupe d'individus appelées profils : ils représentent leurs intérêts à long terme.

4. Recherche d'information versus filtrage d'information

4.1 Fonctionnement de base

La recherche d'information est étroitement liée au filtrage de l'information dans le sens où ils ont le même but qui est de retrouver l'information pertinente pour un certain utilisateur.

La recherche d'information permet de retourner à partir d'un ensemble de documents, ceux dont le contenu correspond le mieux à un besoin en informations d'un utilisateur, exprimé à l'aide d'une requête, d'un point de vue très large, comme étant un processus de sélection de l'information.

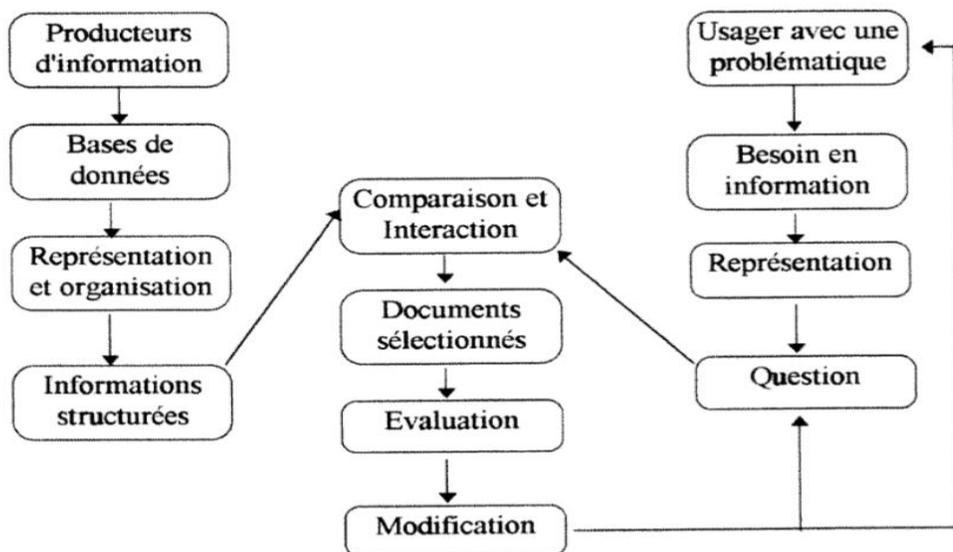


Figure 3 : modèle de recherche d'information

Le filtrage d'information « achemine des documents qui se présentent vers des groupes de personnes, en se basant sur leurs profils à long terme », et élaborés à partir de données d'apprentissage, alors le filtrage de l'information est simplement. Un cas particulier de la recherche d'information où l'information arrive d'une manière dynamique.

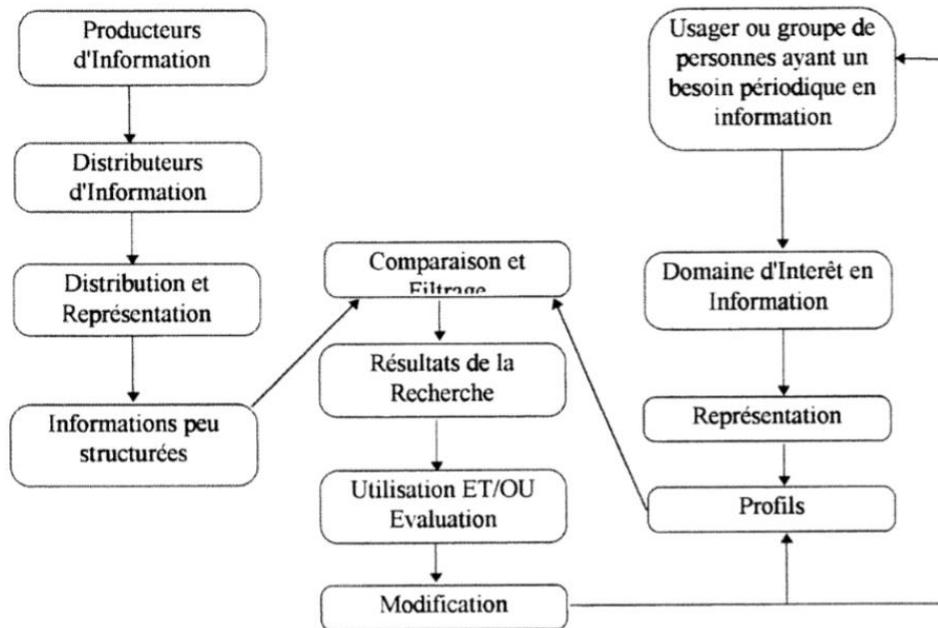


Figure 4 : modèle de filtrage d'information

4.2 Différence entre la recherche et le filtrage

La recherche d'information concerne des usages singuliers du système, avec une personne ayant un objectif et une requête à la fois, alors que le filtrage d'information est concerné par des usages répétitifs du système, par une ou plusieurs personnes avec des buts et des intérêts à long terme.

La recherche d'information se base sur l'adéquation de la requête avec les documents pertinents à sélectionner des collections, en revanche pour le filtrage de l'information la requête est remplacée par le profil de l'utilisateur qui est une sorte de requête à long terme.

La recherche d'information sélectionne des documents à partir d'une base classique de documents, par contre le filtrage d'information filtre ou élimine des documents à partir d'un flux dynamique de données.

La recherche d'information assure la collection et l'organisation des documents, le filtrage de l'information assure la distribution des documents aux personnes qui en ont besoin.

Le filtrage de l'information est généralement appliqué à des flux arrivants de nouvelles données tandis que dans la recherche d'information, les modifications dans la base de données ne sont pas fréquentes et la recherche n'est pas limitée aux nouveaux documents. Le filtrage est basé sur des descriptions des préférences d'un individu ou d'un groupe d'individus appelées profils, Ils représentent leurs intérêts à long terme. Contrairement dans la recherche d'information les requêtes reflètent des intérêts à court terme.

Cette comparaison est résumée dans le tableau suivant :

	Recherche d'information	Filtrage d'information
Approche	- Trouver l'information Recherchée	- Distribuer l'information
Livraison	- Corpus statique, sur demande	- Flux dynamique
Persistance	- Des besoins à court terme	- Des intérêts à long terme
Analyse du contenu	- Utilise souvent des mots-clés	- Différents et multiples dispositifs utilisés
Personnalisation	- Non personnalisé	- Profil d'utilisateur requis
Fonctionnalités	- Non personnalisé - Non adaptatif - Non dynamique - A court terme	- Personnalisé - S'adapte au changement du profil de l'utilisateur - Filtre dynamiquement l'information entrante - A long terme

Tableau 1 : Comparaison entre recherche d'information et filtrage d'information.

5. Types de filtrage d'information

Généralement, il existe trois grandes approches de filtrage : basé sur le contenu, collaboratif et hybride.

5.1 Filtrage basé sur le contenu

5.1.1 Définition

Le filtrage basé sur le contenu est une évolution générale des études sur le filtrage d'information, s'appuie sur le contenu des documents (thèmes abordés) pour les comparer à un profil lui-même constitué de thèmes.

Le premier avantage du filtrage basé sur le contenu est qu'il peut répondre aux intérêts à long terme des utilisateurs, en employant des techniques efficaces dans le domaine de l'intelligence artificielle pour la mise à jour des profils et le recoupement entre profils et documents.

5.1.2 Les limits

La difficulté d'indexation de documents multimédia : Le filtrage basé sur le contenu s'appuie sur un profil qui décrit le besoin de l'utilisateur du point de vue thématique, de façon analogue à une requête qui serait destinée à un système de recherche d'information. Ce profil peut prendre diverses formes, mais il repose toujours sur des termes qui seront comparés aux termes qui indexent le document. De ce fait, la difficulté d'indexer des documents, multimédia ou non, est un goulet d'étranglement pour cette approche.

L'incapacité à traiter d'autres critères de pertinence que les critères strictement thématiques posent également problème. Le filtrage des documents basé sur le contenu ne permet pas d'intégrer d'autres facteurs de pertinence que le facteur thématique. Pourtant il existe de nombreux autres facteurs de pertinence comme par exemple l'adéquation entre le public visé par l'auteur et l'utilisateur, ou encore la qualité scientifique des faits présentés, la fiabilité de la source d'information, le degré de précision des faits présentés, etc.

Enfin l'effet dit « entonnoir » restreint le champ de vision des utilisateurs. En effet, le profil évolue toujours dans le sens d'une expression du besoin de plus en plus spécifique, qui ne

laisse pas de place à des documents pourtant proches mais dont la description thématique diffère fortement.

5.2 Filtrage collaboratif

5.2.1 Définition

Le filtrage collaboratif consiste à filtrer les documents du flux entrant en se basant sur l'opinion que, chaque utilisateur de la communauté a porté dessus. Tout document qu'il l'aura alors jugé intéressant, sera diffusé à l'ensemble des utilisateurs ayant eu des opinions similaires par le passé.

Un système de filtrage collaboratif est constitué de trois principales fonctionnalités : le calcul de la prédiction de l'évaluation du document par l'utilisateur, le calcul de la proximité entre les utilisateurs en plus de la mise à jour périodique des profils utilisateurs à chaque nouvelle évaluation de documents.

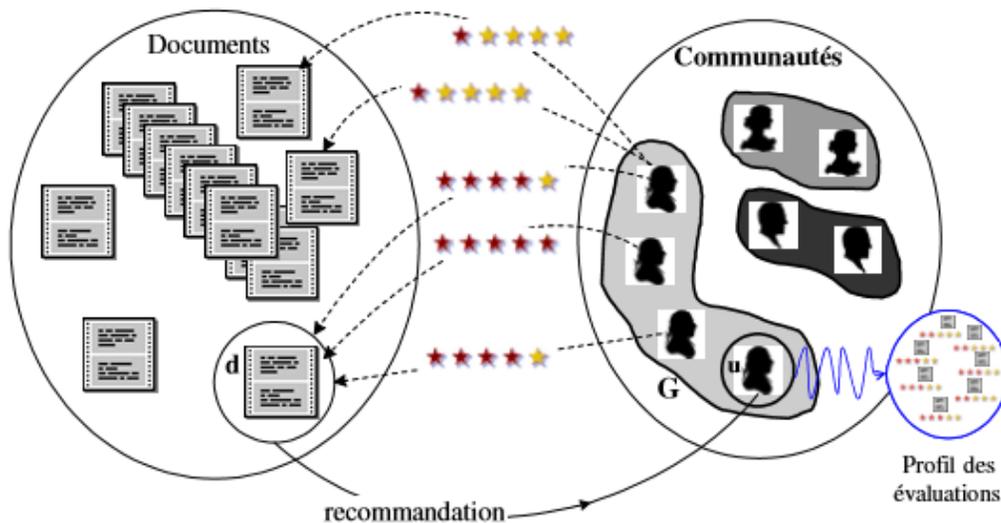


Figure 5 : Principe général de filtrage collaboratif.

Le filtrage collaboratif est parmi les technologies les plus populaires dans le domaine des systèmes de recommandation [1]. Il fonctionne en recommandant des Items en fonction du comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires et cette tâche est faite en utilisant des méthodes qui collectent et analysent des données sur le comportement, les activités, les préférences des utilisateurs.

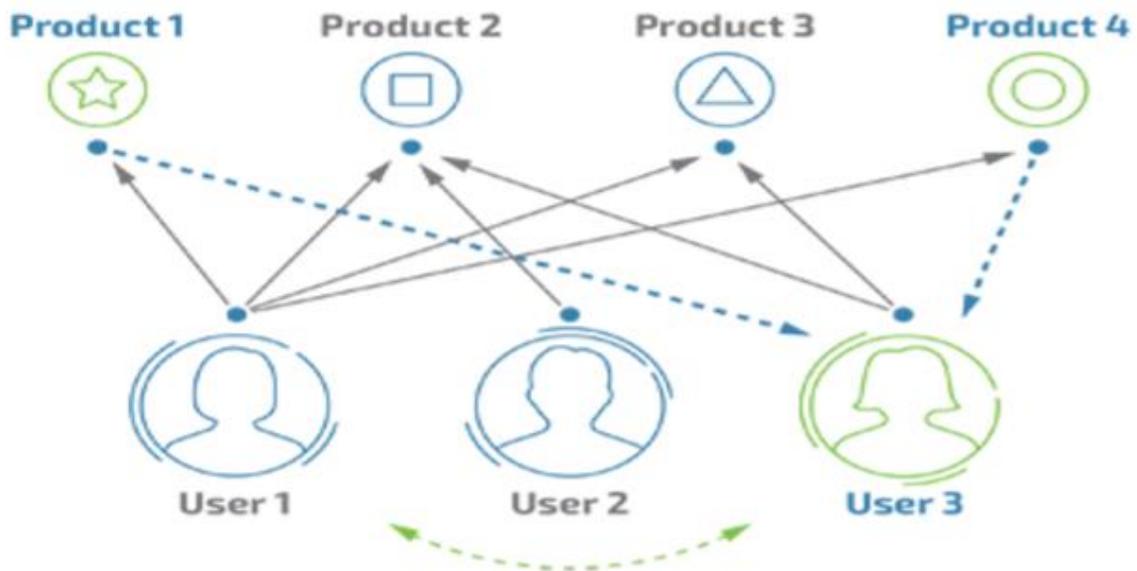


Figure 6 : filtrage collaboratif.

5.2.2 Architecture générale

L'architecture générale d'un système de filtrage collaboratif s'articule autour de deux fonctionnalités centrales : le calcul de la proximité entre les utilisateurs, et le calcul de la prédiction de l'évaluation qu'un utilisateur fera d'un document. S'ajoute la fonctionnalité de mise à jour perpétuelle des profils d'utilisateurs, au fur et à mesure de la collecte de leurs évaluations.

Dans ce type de système, où l'utilisateur contribue de façon décisive au bon fonctionnement du système dans son ensemble, on ne peut négliger les fonctionnalités interactives du système. Les fonctionnalités indispensables sont les suivantes : une interface permettant d'évaluer un document, une interface permettant de visualiser les documents reçus par filtrage.

D'autres fonctionnalités interactives peuvent exister, notamment celle permettant aux utilisateurs d'effectuer ce que l'on appelle du « filtrage actif » : le terme « filtrage actif » [2] traduit le fait que l'utilisateur décide, de sa propre initiative, d'envoyer des documents à certains membres de la communauté. Cette possibilité peut s'avérer très utile lors de l'amorçage du système, pour faire croître les chances de recoupement des profils d'utilisateurs. En effet, un utilisateur qui reçoit un document envoyé par un autre est amené à l'évaluer lui aussi ; à la suite de cette évaluation, son propre profil et celui de l'auteur du filtrage actif se recouperont nécessairement

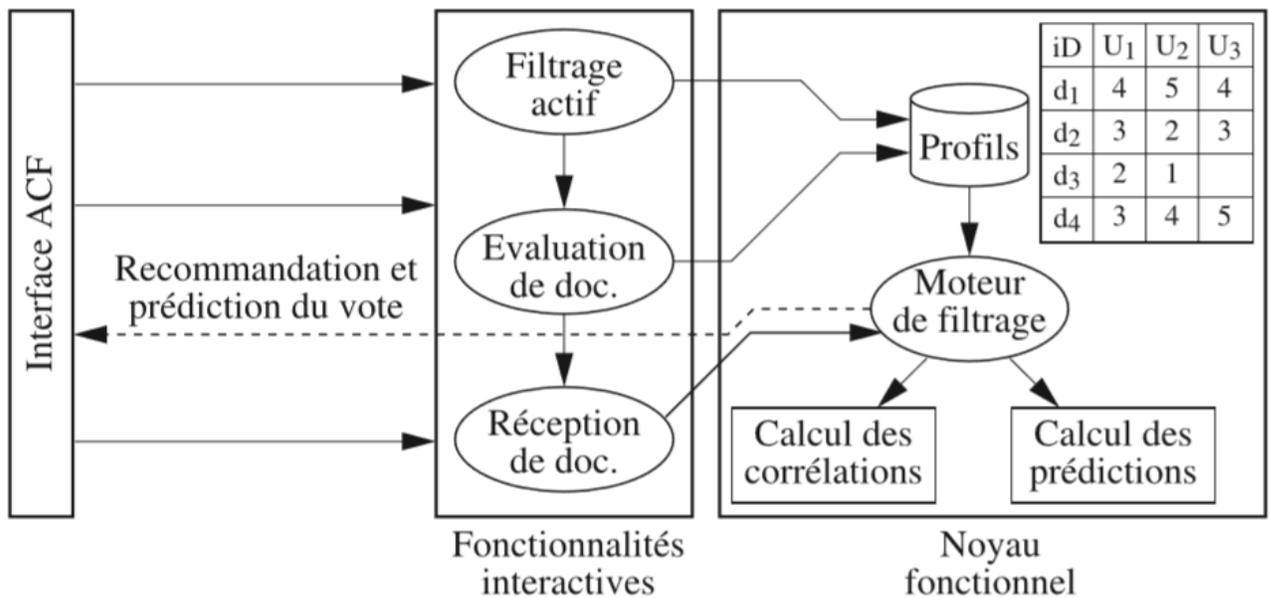


Figure 7 : Architecture générale d'un système de filtrage collaboratif

5.2.3 Calcul de la prédiction pour un système de filtrage collaboratif :

Les systèmes de filtrage collaboratif peut être classés en deux grandes catégories : les algorithmes basés « mémoire », les algorithmes basés « modèle »

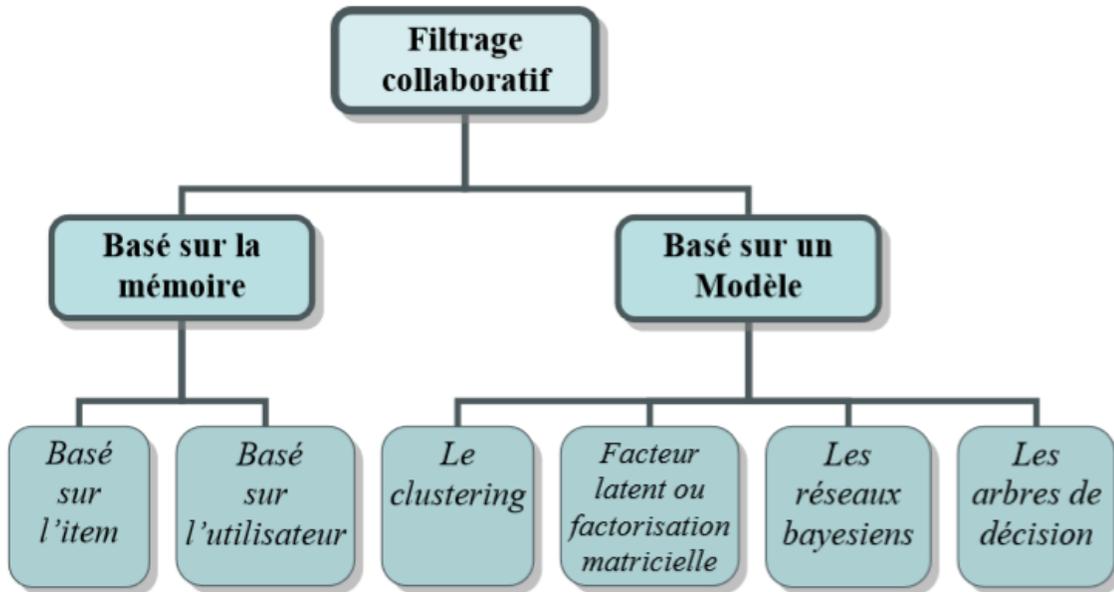


Figure 8 : Les méthodes de filtrage collaboratif.

Les algorithmes basés mémoire utilisent l'ensemble de la base de données des évaluations des utilisateurs pour faire les prédictions : les évaluations de l'utilisateur actif sont prédites à partir d'informations partielles concernant l'utilisateur actif, et un ensemble de poids calculés à partir de la base de données des évaluations des utilisateurs.

Si I_i est l'ensemble des items évalués par l'utilisateur i , alors l'évaluation moyenne pour l'utilisateur i peut être définie comme :

$$v_i = \frac{1}{|I_i|} + \sum_{j \in I_i} v_{i,j} \quad (1.1)$$

L'évaluation prédit sur l'item j pour l'utilisateur actif a est une somme pondérée des évaluations des autres utilisateurs :(2)

$$p_{a,j} = \bar{v}_a + \frac{\sum_{i=1}^n w(a,i)(v_{i,j} - \bar{v}_i)}{\sum_{i=1}^n |w(a,i)|} \quad (1.2)$$

Où n est le nombre d'utilisateurs dans la base de donnée qui ont un poids non nul, et est un facteur de normalisation tel que la somme des valeurs absolues des poids fait 1. Le poids $w(a,i)$ est déterminé de façon variable, selon l'algorithme.

Les détails de calcul de ces poids donnent lieu à des algorithmes différents. Ici la présentation de l'algorithme basé sur la corrélation, et l'algorithme basé sur la similarité de vecteurs.

Pour l'algorithme basé sur la corrélation, le poids est calculé comme la corrélation entre les utilisateurs a et i , comme suit :(3)

$$w(a, i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}} \quad (1.3)$$

Où les sommes sur les j concernent les items pour lesquels à la fois i et a ont donné des évaluations.

Pour l'algorithme basé sur la similarité des vecteurs, le poids est calculé comme un cosinus entre les vecteurs formés par les évaluations des utilisateurs, comme suit :

$$w(a, i) = \frac{\sum_j v_{a,j} v_{i,j}}{\sqrt{\sum_j v_{a,j}^2 \sum_j v_{i,j}^2}} \quad (1.4)$$

L'erreur absolue de chaque note estimée doit être minimisée.

Ainsi, si, pour l'utilisateur actif a , les $v_{a,j}$ sont les notes réelles attribuées aux articles j de l'ensemble test P_a (de taille m_a), et les $p_{a,j}$ sont les valeurs estimées pour les mêmes articles, alors on calcule l'erreur moyenne dans l'estimation des notes comme suit :

$$MoyError(a) = \frac{1}{m_a} \sum_{j \in P_a} |p_{a,j} - v_{a,j}| \quad (1.5)$$

Les algorithmes basés « modèle » utilisent la base de données des évaluations des utilisateurs pour estimer ou apprendre un modèle qui est alors utilisé pour les prédictions.

Du point de vue probabiliste, la tâche de prédiction d'une évaluation peut être vue comme le calcul de la valeur espérée d'une évaluation, étant donné ce que l'on sait d'un utilisateur.

Supposons que les évaluations se fassent sur une échelle d'entiers de 0 à m. Alors la valeur prédite sera :

$$p_{a,j} = E(v_{a,j}) = \sum_{i=0}^m Pr(v_{a,j} = i | v_{a,k}, k \in I_a) i \quad (1.6)$$

Où la probabilité exprimée est celle dont l'utilisateur actif fera l'évaluation particulière i pour l'item j compte tenu des évaluations observées auparavant.

Le calcul préconise que la probabilité exprimée est celle dont l'utilisateur actif fera, compte tenu des évaluations observées auparavant. Deux modèles pouvant être employés pour effectuer les calculs de ces probabilités : le modèle à base de clusters et le modèle à base de réseau bayésien

Le modèle à base de clusters repose sur le principe que certains groupes ou types d'utilisateurs capturent un ensemble commun de préférences et de goûts. Etant donné un tel groupe, les préférences concernant les différents items (sous la forme d'évaluations) sont indépendantes.

Du point de vue formel, on s'appuie sur un classifieur bayésien, où la probabilité des évaluations est conditionnellement indépendante sachant l'appartenance à une variable de classe C non observée comportant un ensemble de valeurs discrètes relativement petit. Le modèle de probabilité qui met en relation les probabilités jointes des classes et des évaluations, et un ensemble de distributions conditionnelles et marginales, est la formulation standard « naïve » de Bayes :

$$Pr(C = c, v_1, \dots, v_n) = Pr(C = c) \prod_{i=1}^n Pr(v_i | C = c) \quad (1.7)$$

La partie gauche est la probabilité d'observer un individu d'une classe particulière et un ensemble complet de valeurs d'évaluations. Dans ce contexte, on peut calculer directement les expressions de probabilité requises pour l'équation précédente donnant la valeur prédite. Les paramètres du modèle, les probabilités d'appartenance à une classe $Pr(C = c)$, et les probabilités conditionnelles des évaluations sachant la classe sont estimées à partir d'un ensemble d'exemples d'évaluations d'utilisateurs, appelé la base des évaluations.

Le modèle à base de réseau bayésien associe un nœud à chaque item du domaine. Les états pour chaque nœud correspondent aux valeurs d'évaluation possibles pour chaque item. On

inclut également un état correspondant à l'absence d'évaluation pour les domaines où il n'y a pas d'interprétation naturelle pour les données manquantes.

On peut alors appliquer un algorithme d'apprentissage de réseau bayésien sur la base d'exemples, où les évaluations manquantes sont associées à une valeur « pas d'évaluation ». L'algorithme d'apprentissage cherche sur plusieurs structures de modèle en termes de dépendances pour chaque item. Dans le réseau résultant de l'apprentissage, chaque item a un ensemble d'items « parent » qui sont les meilleurs prédicteurs de ses évaluations. Chaque table de probabilité conditionnelle est représentée par un arbre de décision qui code les probabilités conditionnelles pour ce nœud.

5.2.4 Systèmes de filtrage collaboratif

Il y a trois processus principaux dans un système de filtrage collaboratif : évaluation des recommandations, formation des communautés et production des recommandations. Parmi ces trois processus, la formation des communautés, ou plus généralement la gestion des communautés.

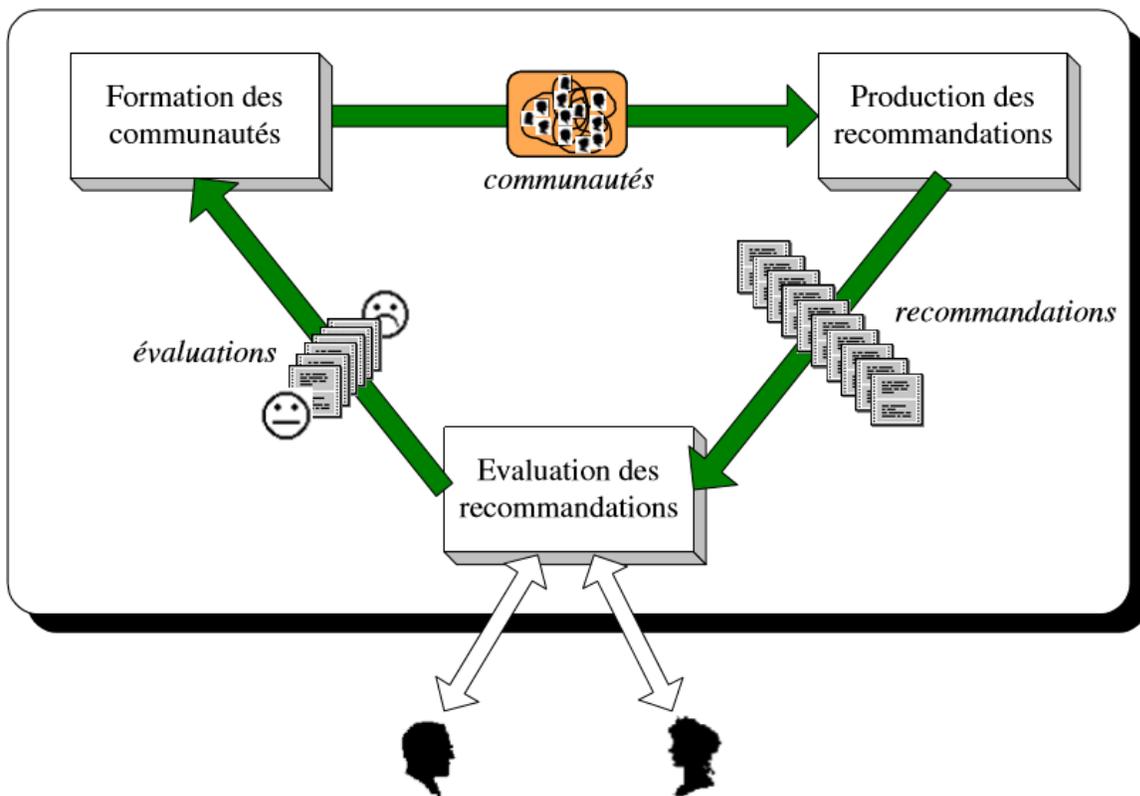


Figure 9 : Trois processus principaux d'un système de filtrage collaboratif.

5.2.4.1 Evaluation des recommandations

Selon le principe de base du filtrage collaboratif, les utilisateurs doivent fournir leurs évaluations sur des documents afin que le système forme les communautés. Évaluer une recommandation peut se faire de façon explicite ou implicite, comme suit.

Explicite : L'utilisateur donne une valeur numérique sur une échelle donnée ou une valeur qualitative de satisfaction, par exemple, mauvaise, moyenne, bonne et excellente.

Implicite : Le système induit la satisfaction de l'utilisateur à travers ses actions. Par exemple, le système estimera qu'une recommandation supprimée correspond à une évaluation très mauvaise, alors qu'une recommandation imprimée ou sauvegardée peut être interprétée comme une bonne évaluation.

5.2.4.2 Formation des communautés

Le processus de formation des communautés est le noyau d'un système de filtrage collaboratif. Pour chaque utilisateur, le système doit calculer sa communauté, généralement cela se fait par la proximité des évaluations des utilisateurs. Comme mentionné précédemment.

5.2.4.3 Production des recommandations

Dans ce dernier processus, une fois la communauté de l'utilisateur créée, le système prédit l'intérêt qu'un document particulier peut présenter pour l'utilisateur en s'appuyant sur les évaluations que les membres de la communauté ont faites de ce même document. Lorsque l'intérêt prédit dépasse un certain seuil, le système recommande le document à l'utilisateur.

5.2.4.4 Profils et communautés

Ici, nous discutons les profils basés sur l'historique des évaluations des utilisateurs, ainsi que les communautés, qui sont les deux facteurs clés d'un système de filtrage collaboratif.

Le profil de l'utilisateur exprime l'ensemble de ses intérêts relativement stables à long terme, intérêt qu'il peut partager avec un groupe ou une communauté d'individus.

Le filtrage collaboratif est réalisé à l'aide d'un profil constitué de l'historique des évaluations de chaque utilisateur. Bien que le filtrage collaboratif puisse s'appliquer sur des évaluations explicites aussi bien que sur des évaluations implicites, dans cette thèse Doit être pris en compte essentiellement les évaluations explicites. Ainsi, un profil est constitué

de paires (document, évaluation), et il s'enrichit progressivement au fur et à mesure que l'utilisateur évalue des documents reçus.

La notion de communauté dans un système de filtrage collaboratif est définie comme le regroupement des utilisateurs en fonction de l'historique de leurs évaluations, afin que le système calcule des recommandations. Selon cette optique, les profils sont un facteur interactif, alors que les communautés sont considérées comme un facteur interne du système.

5.2.5 Le filtrage collaboratif et le basé sur le continu

Le filtrage collaboratif permet de contourner les limitations du filtrage par le contenu :

Le problème de l'indexation des documents multimédia ne se pose plus, du fait que la sélection et le filtrage ne s'opèrent plus par le contenu mais plutôt en fonction des opinions des utilisateurs sur les documents.

Cette « indexation parallèle » présente un autre avantage, En effet, lorsqu'un utilisateur émet une opinion positive sur un document, il affirme que le document traite bien d'un sujet qui l'intéresse, mais aussi que ce document est de bonne qualité, Ainsi le problème de l'incapacité à traiter d'autres critères est également résolu.

Enfin, l'effet « entonnoir » est lui aussi éliminé du fait que les documents entrants ne sont pas filtrés en fonction du contenu. Pour qu'un utilisateur reçoive un document, il suffit qu'un autre utilisateur de profil proche l'ait jugé intéressant, et cela quels que soient les termes qui indexent le contenu du document.

5.2.6 Limites du filtrage collaboratif

Le démarrage à froid : Ce phénomène se produit en début d'utilisation du système, dans des situations critiques où le système manque de données pour procéder à un filtrage personnalisé de bonne qualité, ce qui conduit à l'impossibilité de fournir des recommandations pertinentes.

La masse critique : Afin de former de meilleures communautés, le système exige un nombre suffisant d'évaluations en commun entre les utilisateurs pour les comparer entre

eux. Malgré la taille énorme de l'ensemble des documents, achats, etc. dans les systèmes, le nombre des évaluations en commun entre utilisateurs risque d'être faible.

Rapport coût-bénéfice : Le filtrage collaboratif est un processus qui implique fortement les utilisateurs, ils doivent chacun émettre suffisamment d'évaluations pour dépasser le problème du démarrage à froid. En effet, l'utilisateur ne perçoit pas toujours favorablement le rapport coût-bénéfice que ce type de système apporte. Lorsqu'il évalue des documents, l'utilisateur se demande si ses efforts seront payés en retour à court ou à moyen terme.

L'expression limitée du besoin : En général, les utilisateurs ne peuvent exprimer l'évolution de leur besoin d'information que sous la forme d'une succession d'évaluations des documents reçus. Ainsi, un changement dans le besoin d'information sera potentiellement mal traduit par un utilisateur qui ne reçoit pas du système les documents lui permettant d'exprimer ce changement.

5.3 Filtrage Hybride

Constatant les avantages et inconvénients de chacune des deux approches ci-dessus, on comprend que de nombreux systèmes reposent sur leur combinaison, ce qui en fait des systèmes de filtrage dits « hybrides ».

Cette approche est en fait une combinaison des deux types d'approches précédemment cités [4]. Les systèmes hybrides tirent parti des avantages des deux approches citées précédemment.

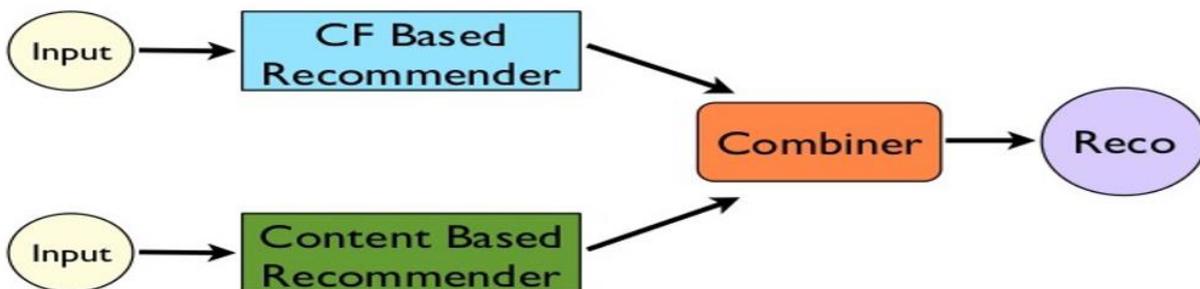


Figure 10 : filtrage hybride.

6. Conclusion

Dans ce chapitre, nous avons donné une définition de filtrage d'information et la différence entre le filtrage et la recherche, en citant les différentes approches les plus utilisées, et nous avons approfondi l'explication du filtrage collaboratif.

De cette manière, certains problèmes se posent, par exemple, si une grande matrice d'évaluation est utilisée, le problème lié au grand fonctionnement du compte est que les périphériques normaux sont incapables d'effectuer des opérations. Dans ce cas, les développeurs décident d'utiliser d'autres techniques. Appelées informatique en nuage (Cloud computing).

Consiste à exploiter la puissance de calcul ou de stockage de serveurs informatiques distants par l'intermédiaire d'un réseau, généralement Internet. Nous allons le détailler dans le chapitre 2.

Chapitre n°2 : services du cloud computing.

1. Introduction

Indéniablement, la technologie de l'internet se développe d'une manière exponentielle depuis sa création. Actuellement, une nouvelle "tendance" a fait son apparition dans le monde de l'IT (information Technologies : Technologies de l'information et de la communication), il s'agit du Cloud Computing. Cette technologie, s'appuie sur le WEB, offre des occasions aux sociétés de réduire les coûts d'exploitation des logiciels par leurs utilisations directement en ligne [9].

L'Internet actuel fournit un contenu sous les formes des vidéos, e-mails et les informations servis dans les pages web. Avec le Cloud Computing, la prochaine génération d'internet va nous permettre d'acheter des services informatiques. Nous serions en mesure de louer à partir d'une vitrine virtuelle la base nécessaire pour construire un centre des données virtuel, tels que le processeur, la mémoire et le stockage, et ajouter au-dessus le middleware nécessaire tels que les serveurs d'applications Web, bases de données et bus serveur d'entreprise, etc. [11]

Ce chapitre, présente globalement des méthodes pour passer à l'échelle par le système de multithreading et le Cloud computing, et il précise sur le cloud, suivi d'une définition de ce dernier qui sera basée sur une analyse des définitions proposées par le monde académique. Ils décrivent aussi les services de Cloud, les types de Cloud et ses acteurs ainsi que les avantages, les inconvénients, les objectifs principaux du Cloud computing, avec une petite présentation sur le multithreading et la différence entre eux et le multiprocesseur.

2. Passage à l'échelle en filtrage d'information (Scalability)

La quantité d'information numérique produite et consultée a considérablement augmenté et sa diversité s'est accrue. Or, le facteur d'échelle joue un rôle important dans la quantité et la qualité des traitements que l'on peut appliquer aux informations, aussi bien de manière intrinsèque que perçue par l'utilisateur. Propose un panorama des problèmes qui découlent de ces évolutions ainsi que quelques pistes de recherche afin de répondre à ce qui semble bien être un nouveau défi de « passage à l'échelle » [17].

Le traitement de grands volumes d'informations est souvent désigné par l'expression « passage à l'échelle ». Plus précisément, le passage à l'échelle d'une technique ou d'un algorithme désigne sa capacité à traiter des volumes considérables d'informations tout en conservant une complexité du même ordre de grandeur réelle, c'est-à-dire chronométrée, que celle induite par le traitement des volumes antérieurs moins importants. Lorsque l'on envisage l'accès à de grandes masses d'informations, il faut également traiter de la

présentation des résultats. On peut alors légitimement se demander si le passage à l'échelle influence les modes de présentations des données. De manière générale, les travaux sur la visualisation de grandes masses de données ont été réalisés principalement dans le domaine des interfaces homme-machine même s'il y a des travaux spécifiques aux systèmes de filtrage d'information. Les problèmes du passage à l'échelle dans la taille des corpus dans un système de filtrage d'information rendent plus difficile l'évaluation et donc la perception des bonnes réponses dans un espace devenu plus vaste. Les études liées à la présentation de grandes masses de données doivent donc s'adapter au filtrage d'information.

On peut considérer, que le problème du passage à l'échelle est seulement un problème de volume. Or, ce problème entraîne un problème de temps de calcul

Les propositions de réduction de l'échelle ont porté essentiellement sur utilisation des API du cloud computing.

Comme nous discutons précédemment, le filtre à une approche, et chacun a une algorithmes, chaque algorithmes a une complexité. Donc on discutons sur ce approche.

Les algorithmes basés mémoire : ces algorithmes se compose à deux approches basées sur l'utilisateur et basées sur items. Et chaque approche a une complexité.

Pour l'approche basées sur l'utilisateur un voisinage pour chaque utilisateur peut être considéré. Dans un tel cas, la taille de voisinage K est alors un paramètre système à définir, et seuls les voisins de l'utilisateur actif sont pris en compte pour les prédictions. La complexité temporelle des approches basées sur l'utilisateur est $O(N \times M \times K)$ pour la construction du modèle, $O(K)$ pour une prévision d'évaluation. En ce qui concerne les approches basées sur l'utilisateur, un voisinage pour chaque élément peut être considéré. Dans un tel cas, la taille du voisinage K est alors un paramètre système à définir, et seuls les voisins de l'élément cible sont pris en compte pour les prédictions.

La complexité des approches basées sur les items est $O(M \times N \times K)$ pour la construction du modèle.

3. Cloud computing

3.1. Définition

Le Cloud Computing est un concept qui consiste à accéder à des données et des services sur un serveur distant. Traditionnellement, une entreprise utilisait sa propre infrastructure pour héberger ses services. Elle achetait donc ses propres serveurs, et assurait le développement et la maintenance des systèmes nécessaires à son fonctionnement. Par opposition, le Cloud Computing se repose sur une architecture distante, Le fournisseur donc

assure la continuité du service et de la maintenance. Les services de Cloud Computing sont accessibles via un navigateur web [6].

3.2. Les caractéristiques essentielles du Cloud computing

Le modèle Cloud Computing se différencie par les cinq caractéristiques essentielles suivantes [7] :

- **Accès aux services par l'utilisateur à la demande**

La mise en œuvre des systèmes est entièrement automatisée et c'est l'utilisateur, au moyen d'une console de commande, qui met en place et gère la configuration à distance.

- **Accès réseau large bande**

L'ensemble des ressources doit être accessible et à disposition de l'utilisateur universellement et simplement à travers le réseau, quels que soient les clients utilisés (serveur, PC, client mobile, etc.). Ceci implique trois prédispositions :

Accès utilisateur : afin de garantir que l'accès à une solution de cloud soit perçu comme stable et fiable, il est nécessaire que celle-ci dispose d'un accès haut débit

Accès applicatif : en vue de s'assurer que la solution soit en mesure de mettre à disposition de l'utilisateur les autres fonctionnalités du cloud, l'accès au réseau doit être disponible au sein même de la solution

Accès opérateur : dans le but de pouvoir maintenir et administrer un système cloud correctement, le fournisseur de service cloud doit pouvoir accéder à celui-ci via le réseau

Ainsi, s'assurer qu'un accès total au réseau est disponible dans tous les aspects d'une solution de cloud est essentiel pour offrir l'ensemble des autres caractéristiques.

- **Réservoir de ressources (non localisées)**

Les ressources matérielles du fournisseur sont partagées entre les différents utilisateurs du service. Ce partage rend l'emplacement exact des données des utilisateurs impossible à déterminer, ce qui peut poser un véritable problème aux entreprises qui sont soumises à de nombreuses contraintes réglementaires concernant la localisation et le contrôle des

données. En contrepartie, les économies d'échelle réalisées permettent de réduire les coûts du fournisseur et donc les dépenses des utilisateurs.

- **Redimensionnement rapide (élasticité)**

L'une des caractéristiques du cloud computing est l'élasticité des ressources. Cette caractéristique permet aux utilisateurs de provisionner rapidement de nouvelles ressources de manière à être en mesure de répondre à une montée ou à une descente en charge soudaine. Il n'est jamais évident de prévoir les ressources qui seront nécessaires à la mise en place d'un service informatique quelconque, en particulier lorsque ce besoin est en constante évolution. Le cloud computing offre ainsi un moyen de fournir les ressources informatiques nécessaires à une évolution ou à un pic d'utilisation de ce service.

- **Facturation à l'usage**

Les systèmes cloud doivent être capables de s'autocontrôler et de se gérer pour permettre l'optimisation interne du système. Pour cela, ils s'appuient sur des mesures de référence obtenues grâce à divers mécanismes de supervision. Ces mesures précises permettent une juste facturation des utilisateurs ; ceux-ci ne payeront ainsi que pour les ressources qu'ils ont utilisées et seulement pour la durée qu'ils les ont utilisées.



Figure 11 : caractéristique du cloud computing

3.3. Centre de données (Datacenter)

Un centre de traitement de données (data center en anglais) est un site physique sur lequel se trouvent regroupés des équipements constituant le système d'information de l'entreprise (mainframes, serveurs, baies de stockage, équipements réseaux et de télécommunications, etc.). Il peut être interne et/ou externe à l'entreprise, exploité ou non avec le soutien des prestataires. Il comprend en général un contrôle sur l'environnement (climatisation, système de prévention contre l'incendie, etc.), une alimentation d'urgence et redondante, ainsi qu'une sécurité physique élevée. Cette infrastructure peut être propre à une entreprise et utilisée par elle seule ou à des fins commerciales. Ainsi, des particuliers ou des entreprises peuvent venir y stocker leurs données suivant des modalités bien définies [8].

3.4. Types de Cloud Computing

Le concept de Cloud Computing est encore en évolution. On peut toutefois citer trois Types de Cloud Computing [9] :

Le Cloud privé (ou interne) : c'est un réseau informatique propriétaire ou un centre de données qui fournit des services hébergés pour un nombre limité d'utilisateurs.

Le Cloud public (ou externe) : C'est un prestataire de services qui propose des services de stockage et d'applications Web pour le grand public. Ces services peuvent être gratuits ou payants.

Le Cloud hybride (interne et externe) : C'est un environnement composé de multiples prestataires internes et externes.

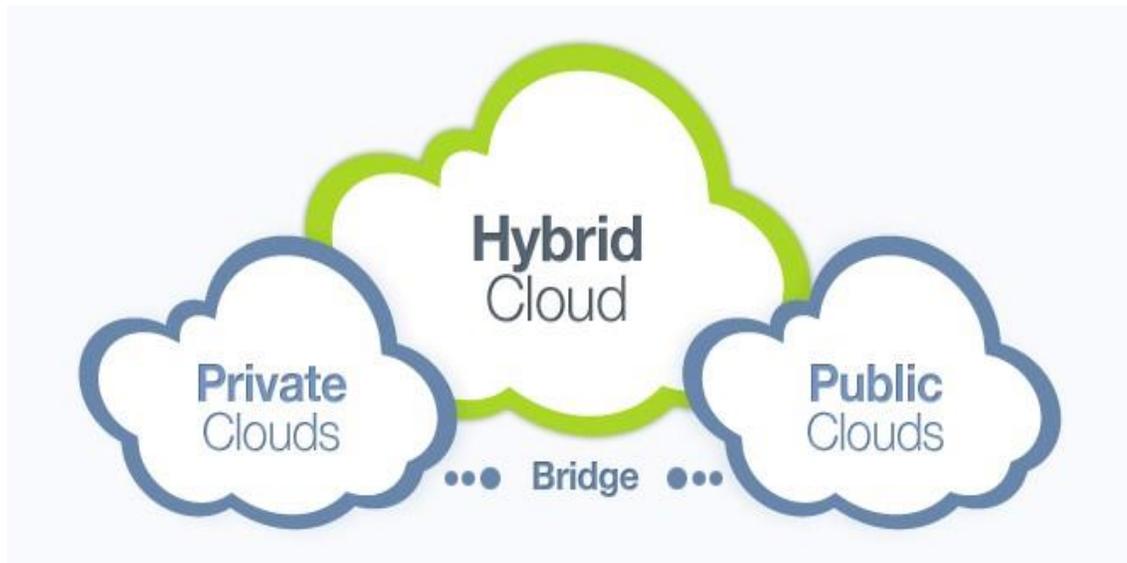


Figure 12 : Types de cloud computing

3.5. Niveaux de service du Cloud Computing

Trois modèles de services peuvent être offerts sur le Cloud : **Software as a Service (SaaS)**, **Platform as a Service (PaaS)**, **Infrastructure as a Service (IaaS)**. Ces trois modèles de service doivent être déployés sur des Infrastructures qui possèdent les cinq caractéristiques essentielles citées plus haut pour être considérées comme du Cloud Computing.



Figure 13 : services du cloud

3.5.1. Le logiciel en tant que service (SaaS)

Est comme son nom l'indique, un modèle de fourniture de logiciels hébergés à distance. L'utilisateur ne gère ni l'infrastructure du cloud ni la plate-forme où l'application s'exécute. Plus besoin d'installer l'application sur ses propres ordinateurs, le client y accède via sa connexion Internet et n'a donc pas à mettre à jour ou à gérer le fonctionnement et la sécurité du logiciel, toutes ces tâches sont effectuées par l'éditeur (le fournisseur). Ce qui simplifie la maintenance et le support. Ces applications sont accessibles à partir de différents périphériques clients par le biais d'une interface client légère, comme un navigateur Web (par exemple : le courrier électronique basé sur le Web), ou une interface spéciale. Parmi les exemples les plus connus, on retrouve : Google Apps, Microsoft Office et OnLive.

3.5.2. Plateforme en tant que service (PaaS)

Le PAAS qui signifie « Platform as a Service » est une architecture composée de tous les éléments nécessaires pour soutenir la construction, la livraison, le déploiement et le cycle de vie complet des applications et des services exclusivement disponibles à partir d'internet. Le PAAS offre des facilités à gérer le déroulement des opérations lors de la conception, du développement, du test, du déploiement et de l'hébergement d'applications web à travers des outils et des services tels que [10] :

- Le travail collaboratif (« team collaboration »).
- L'intégration des services web et bases de données.

Ces services sont fournis au travers une solution complète destinée aux développeurs et disponible immédiatement via l'internet.

3.5.3. Infrastructure en tant que service (IaaS)

L'IAAS (Infrastructure as a Service) est un modèle qui permet de fournir des infrastructures informatiques en tant que service. Ce terme était originellement connu sous le nom de (Hardware as a Service). Ces infrastructures virtuelles composent un des domaines du « As a Service » en empruntant la même philosophie de fonctionnement et de tarification que la plupart des services du Cloud Computing [10].

Plutôt que d'acheter des serveurs, des logiciels, et l'espace dans un centre de traitement de données et/ou de l'équipement réseau, les clients n'ont plus qu'à louer les ressources auprès des prestataires de service. Le service est alors typiquement tarifé en fonction de l'utilisation et de la quantité des ressources consommées.

De ce fait, le coût reflète typiquement le niveau d'activité de chaque client. C'est une évolution de l'hébergement Internet qui se différencie des anciens modes de fonctionnement, on distingue [10] :

Hébergement mutualité : une machine pour plusieurs clients, gérée par un prestataire de service et dont les clients payent le même prix peu importe leur utilisation.

Hébergement dédié : une machine pour un client, gérée le plus souvent par le client lui-même et pour laquelle le client paye le même prix chaque mois peu importe son utilisation.

Infrastructure as a Service : un nombre indéfini de machines pour un nombre indéfini de clients, dont les ressources sont combinées et partagées pour tous les clients. Chaque client paye en fonction de son utilisation de l'architecture.

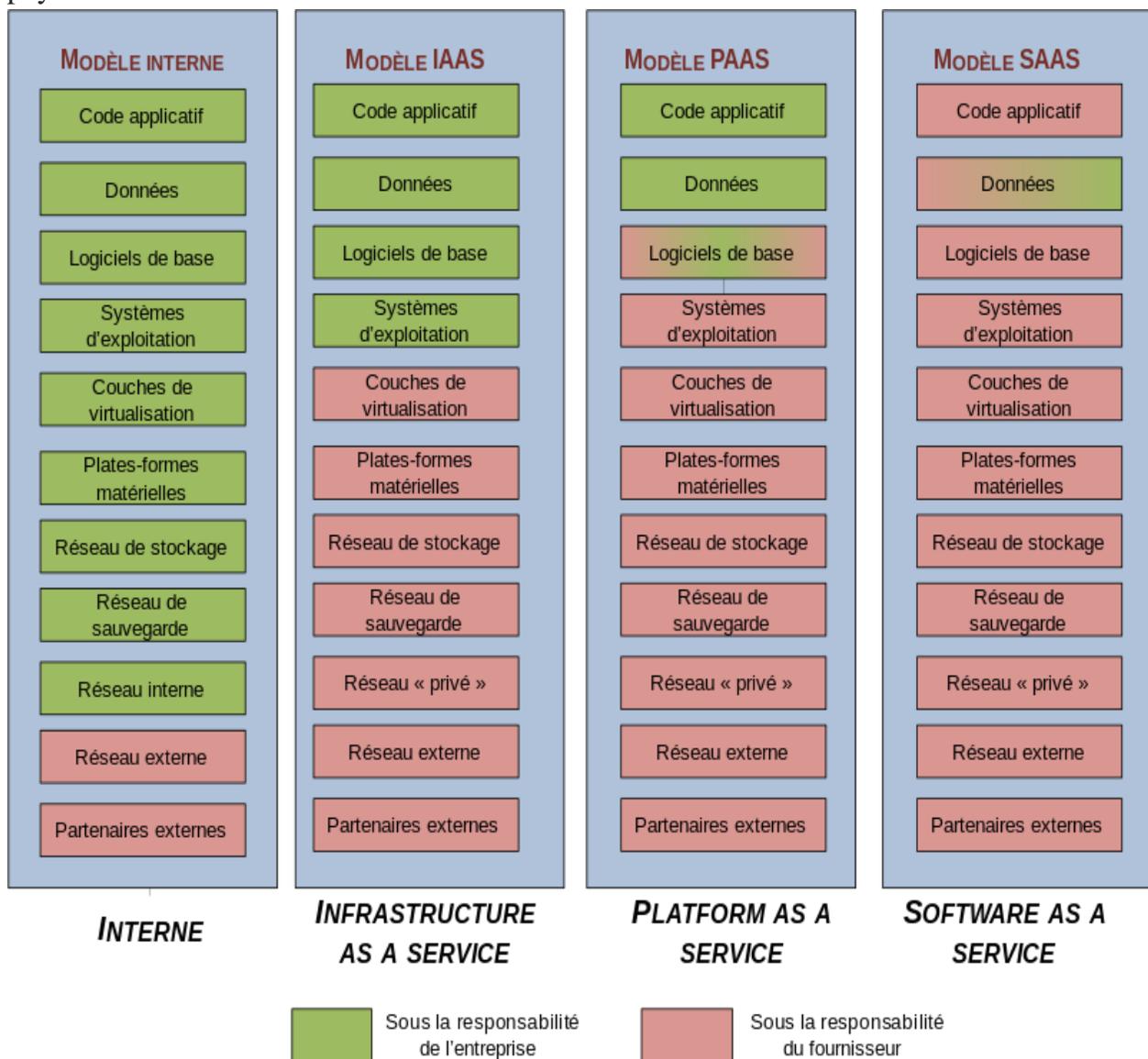


Figure 14 : Niveau de service cloud computing.

3.6. Avantages et inconvénients des services

Du point de vue économique, le Cloud Computing est essentiellement une offre commerciale d'abonnement économique à des services externes. Selon le National Institute of Standards and Technology, il existe trois catégories de services qui peuvent être offerts en Cloud Computing : IaaS, PaaS et SaaS [12]. Les avantages et les inconvénients de ces services se résume dans le tableau ci-dessous.

	Avantage	Inconvénient
SaaS	<ul style="list-style-type: none"> - Pas d'installation. - Plus de licence. - Migration. - Accessible via un abonnement. 	<ul style="list-style-type: none"> - Logiciel limité. - Sécurité. - Dépendance des prestataires.
PaaS	<ul style="list-style-type: none"> - Pas d'infrastructure nécessaire. - Pas d'installation. - Environnement hétérogène. 	<ul style="list-style-type: none"> - Limitation des langages. - Pas de personnalisation dans la configuration des machines virtuelles.
IaaS	<ul style="list-style-type: none"> - Administration. - Personnalisation. - Flexibilité d'utilisation. - Capacité de stockage infini. 	<ul style="list-style-type: none"> - Sécurité. - Besoin d'un administrateur système. - Demande pour les acteurs du Cloud des investissements très élevés.

Tableau 2 : Avantages et inconvénients des services [9] [13].

3.7. Modèle d'application de Cloud Computing

Le Cloud Computing est un nouveau modèle pour fournir aux entreprises les services informatiques. Ce modèle est basé sur une architecture standard qui contient trois phases : la phase stratégique, la phase de planification et la phase de déploiement, chaque phase contient plusieurs étapes. La figure ci-dessous montre la structure générale de ce modèle [11].

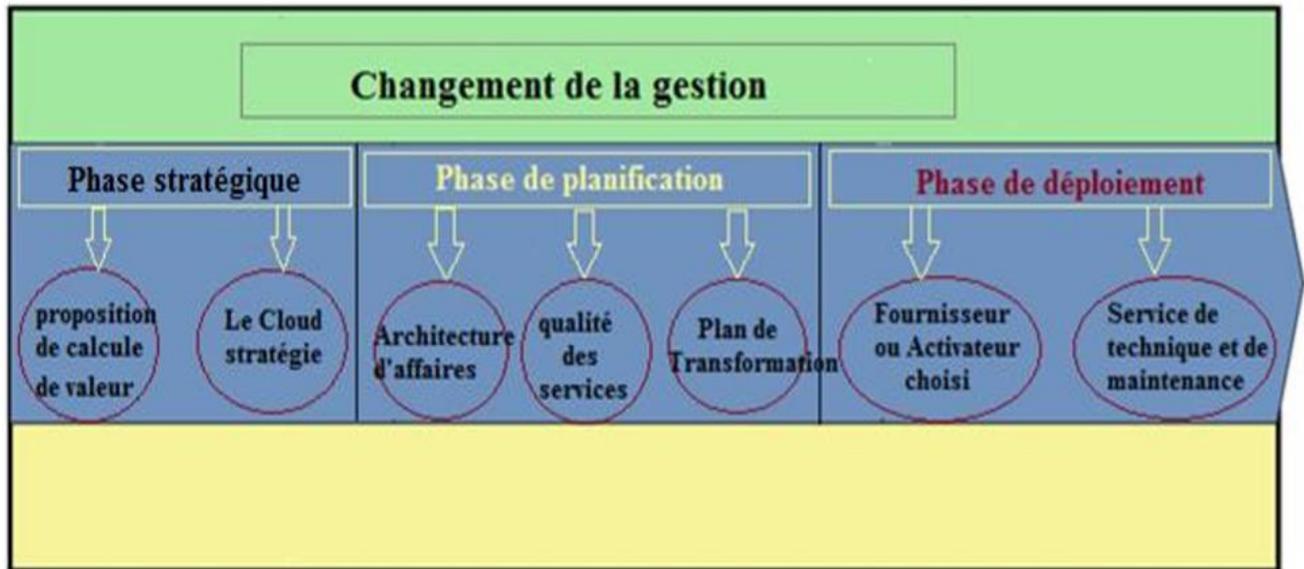


Figure 15 : Modèle d'application de cloud computing.

3.7.1. Phase stratégique

La phase stratégique de Cloud Computing contient deux étapes pour assurer une analyse complète des problèmes, les deux étapes sont : la proposition de calcul de valeur, et le Cloud stratégique.

Proposition de calcul de valeur : L'objectif de cette étape est d'analyser la valeur spécifique de l'entreprise. Dans cette analyse, nous avons besoin d'identifier la cible principale pour le client et appliquer le mode Cloud Computing, et les principaux problèmes qu'elles cherchent à résoudre. Par exemple voici quelques objectifs communs :

- La simplification de la gestion.
- L'exploitation et la réduction des coûts de maintenance
- L'innovation en mode entreprise.
- Faible coût de service d'hébergement.
- Haute qualité de service d'hébergement, etc.

Cloud stratégie : Cette étape est la partie la plus importante de la phase stratégique. La mise en place de Cloud stratégie est basée sur le résultat d'analyse de l'étape proposition de calcul de valeur. L'analyse professionnelle faite par cette étape implique généralement une large clientèle d'affaires de modèle de recherche, de l'organisation d'analyse, de la structure et de processus d'identification d'opération, aussi, il y'a la limitation dans le plan, telles que le souci de norme de sécurité, l'exigence de fiabilité et les règles et les règlements.

Basé sur le Cloud stratégie et la proposition de calcul de valeur, ces deux étapes vont analyser le modèle avec une condition nécessaire pour atteindre l'objectif des clients, puis vont mettre en place une stratégie visant la fonction de la ligne directrice.

3.7.2. Phase de planification

Lors de la phase de planification de Cloud Computing, il est nécessaire de connaître les positions des clients et analyser les problèmes et les risques dans l'application de Cloud Computing.

Après cela, des approches concrètes et des plans peuvent être amenés à veiller à ce que les clients peuvent utiliser le Cloud Computing avec succès pour atteindre leurs objectifs d'affaires. Cette phase comprend trois étapes énumérées comme suit :

- Développement d'Architecture d'affaires.
- Développement de qualité des services.
- Développement du Plan de Transformation.

3.7.3. Phase de déploiement

La phase de déploiement se concentre principalement sur la phase stratégique et la phase de planification. Deux étapes sont mises en évidence dans cette phase [11] :

Fournisseur de Cloud Computing ou activateur choisi : Selon les deux phases précédentes, les clients peuvent avoir à choisir un fournisseur de Cloud Computing ou un activateur.

Service de technique et de maintenance : Dans cette étape les différents niveaux de normes sont adoptés ; ces normes sont définies par l'exigence de qualité des services faite au préalable.

3.8. Avantages du Cloud Computing

Un démarrage rapide : Le Cloud Computing permet de tester le business plan rapidement, à coûts réduits et avec facilité.

L'agilité pour l'entreprise : Résolution des problèmes de gestion informatique simplement sans avoir à s'engager à long terme.

Un développement plus rapide des produits : Réduction du temps de recherche pour les développeurs sur le paramétrage des applications.

Pas de dépenses de capital : Plus besoin des locaux pour élargir les infrastructures informatiques [9].

Simplicité : l'entreprise cliente n'a plus besoin de développements coûteux et déplace la responsabilité du fonctionnement du service sur le fournisseur.

3.9. Migration des données

Le chemin vers le cloud est un parcours vers les résultats métier. Le Cloud a aidé des centaines de clients à atteindre leurs objectifs métier à chacune des étapes de leur parcours. Même si le chemin de chaque organisation est unique, il existe des modèles, des approches et des bonnes pratiques communes qui peuvent être implémentés pour rationaliser le processus [14].

Définissez votre approche du cloud computing depuis l'analyse de rentabilité jusqu'à la stratégie de gestion des modifications, en passant par la technologie.

Développez une solide fondation pour les charges de travail de votre entreprise sur cloud en évaluant et en validant votre portefeuille d'applications, et en intégrant votre environnement informatique spécifique aux solutions basées sur les services du cloud.

Concevez et optimisez vos applications métier pour qu'elles offrent une prise en charge du cloud en bénéficiant directement des services cloud.

Répondez à vos exigences de conformité internes et externes en développant et en implémentant des stratégies de sécurité automatisées et des contrôles basés sur des conceptions éprouvées et validées.

Une planification, une communication et une adhésion précoces sont essentielles. La maîtrise de la fonction de forçage (durée, coût, disponibilité, etc.) est essentielle et sera différente pour chaque organisation. Lors de la définition du modèle de migration, les organisations doivent avoir une stratégie claire, élaborer une chronologie du projet réaliste et limiter le nombre de variables et de dépendances pour la transition des applications locales vers le cloud. Tout le long du projet, maintenez l'élan à l'aide des composants clés en organisant des réunions régulières et en rapportant l'avancement et l'état du projet de migration afin que les personnes conservent leur enthousiasme, tout en définissant aussi des attentes réalistes quant à la durée de disponibilité.

Le Cloud Computing a connu un succès rapide notamment du fait qu'il permette aux clients de réaliser une économie d'échelle importante en payant uniquement pour les ressources utilisées tout en déléguant la gestion de l'infrastructure au fournisseur. Cependant, il existe encore un nombre important de défis inhérents à l'adoption du Cloud Computing [15].

Sécurité : Lors de l'adoption du Cloud, le client s'attend à recevoir une sécurité identique, voire meilleure, que celle qu'il aurait eu dans une installation privée. En effet, le fournisseur dispose souvent de plus de moyens pour gérer la sécurité physique de l'infrastructure Cloud, mais aussi celle des données et des applications qui s'y exécutent. Cependant, le fait de déléguer cette sécurité à un tiers comporte un certain nombre de nuances qui doivent être prises en compte par le client.

Tolérance aux fautes et disponibilité : Le fournisseur Cloud est censé offrir à ses clients un environnement tolérant aux fautes ou le client ne risque ni de perdre ses données, ni de voir l'exécution de ses applications perturbées. De même, une certaine disponibilité doit être garantie pour les services Cloud. Ainsi, l'exécution des applications déployées dans le Cloud doit être continue, et le client doit avoir à tout moment la possibilité de déployer une nouvelle application, ou de tout simplement consulter sa consommation en ressources. Ces deux besoins sont souvent bien définis dans un contrat de niveau de service entre le client et le fournisseur. Toute violation à ce contrat peut alors être sanctionnée financièrement ou juridiquement.

L'équilibrage de charge : L'utilisation efficace des ressources et l'équilibrage de charge sont des objectifs ultimes pour les fournisseurs de services, afin de maximiser leurs profits, et assurer une qualité de services optimale [16].

3.10. Fournisseur de services cloud

Le Cloud Computing est la fourniture de ressources informatiques à la demande (puissance de l'ordinateur, stockage de bases de données, applications ainsi que d'autres ressources informatiques) via Internet. Ceci est réalisé en utilisant un réseau de serveurs distants hébergés sur Internet.

Voilà les principaux fournisseurs de services cloud actuelles [19] :

3.10.1. Amazon Web Services (AWS)

Il s'agit d'une plate-forme de cloud computing complète et évolutive fournie par Amazon. AWS propose un mélange de solutions IaaS, PaaS et SaaS. Aws permet à ses abonnés



de profiter d'un cluster d'ordinateurs virtuels à part entière, à tout moment et en fonction de leurs besoins. L'ensemble du service est activé via Internet.

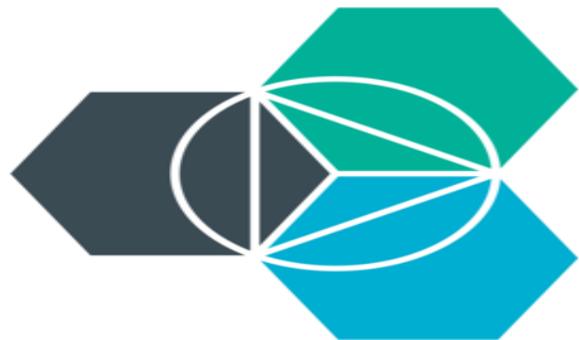
3.10.2. Google Cloud

Ou Google Cloud Platform est la suite de services d'informatique en nuage publics de Google. Il fonctionne sur la même infrastructure que Google utilise en interne pour ses produits destinés aux utilisateurs finaux, tels que Google Search et YouTube. Cela inclut une gamme de services hébergés pour le calcul, le stockage et le développement d'applications exécutés sur le matériel Google. Les services de Google Cloud Platform sont accessibles aux développeurs de logiciels, aux administrateurs de cloud et à d'autres professionnels de l'informatique d'entreprise via Internet public ou via une connexion réseau dédiée.



3.10.3. IBM Bluemix

C'est la solution de cloud computing d'IBM fournie avec la plate-forme (PaaS) ainsi que les offres d'infrastructure en tant que service (IaaS). Avec Bluemix IaaS, les utilisateurs peuvent déployer et accéder à la puissance de calcul, au stockage et au réseau virtualisés via Internet. Les offres de services IBM peuvent être utilisées en tant que modèle public, privé ou hybride, selon les besoins de l'entreprise.



3.10.4. Alibaba Cloud

Filiale du groupe Alibaba, qui fournit une suite de services d'informatique en nuage couvrant l'informatique élastique, le stockage d'objets, la base de données relationnelle, l'analyse de données volumineuses et l'intelligence artificielle dans quinze régions géographiques.



3.10.5. Oracle Cloud

Offre de services cloud d'Oracle Corporation. Ceci fournit des serveurs, un stockage, un réseau, des applications et des services via un réseau mondial de centres de données gérés Oracle. Ici, les services sont fournis sur Internet, sur demande.



3.10.6. Microsoft Azure

(Anciennement Windows Azure) est le service de cloud computing de Microsoft. Activé principalement via des centres de données gérés par Microsoft, ce service s'avère être une solution fiable, en particulier pour les évangélistes Microsoft. Comme les solutions ci-dessus, elle prend en charge le développement, les tests, le déploiement et la gestion d'applications et de services.





Figure 16 : fournisseur de cloud

3.11. Limites du Cloud Computing

Hormis les avantages cités ci-dessus, Cloud Computing possède quelques inconvénients, parmi lesquels [16] :

La bande passante peut faire exploser le budget : La bande passante qui serait nécessaire pour mettre cela dans le Cloud est gigantesque, et les coûts seraient tellement importants qu'il est plus avantageux d'acheter le stockage soi-même plutôt que de payer quelqu'un d'autre pour s'en charger.

Les performances des applications peuvent être amoindries : Un Cloud public n'améliorera définitivement pas les performances des applications.

La fiabilité du Cloud : Un grand risque lorsqu'on met une application qui donne des avantages compétitifs ou qui contient des informations clients dans le Cloud.

Taille de l'entreprise : Si votre entreprise est grande alors vos ressources sont grandes, ce qui inclut une grande consommation du cloud. Vous trouverez peut-être plus d'intérêt à mettre au point votre propre Cloud plutôt que d'en utiliser un externalisé. Les gains sont bien plus importants quand on passe d'une petite consommation de ressources à une consommation plus importante.

4. Multithreading

4.1. Définition

Le multithreading est un type de modèle d'exécution qui permet à plusieurs threads d'exister dans le contexte d'un processus, de sorte qu'ils s'exécutent indépendamment tout en partageant leurs ressources de processus. Un thread conserve une liste d'informations pertinentes pour son exécution, y compris la planification des priorités, les gestionnaires d'exceptions, un ensemble de registres de CPU et l'état de la pile dans l'espace d'adressage de son processus d'hébergement. Le multithread permet de faire des calculs parallèlement si possibles. Ainsi, le résultat souhaité peut être obtenu dans une tranche de temps plus petite.

4.2. Avantage du multithreading

Dans certains cas, les programmes utilisant des threads sont plus rapides que des programmes architectures plus classiquement, en particulier sur les machines comportant plusieurs processeurs. Hormis le problème du coût de la commutation du contexte, le principal surcoût dû à l'utilisation de processus multiples provient de la communication entre processus séparés. En effet, le partage de ressources entre threads permet une communication plus efficace entre les différents *threads* d'un processus qu'entre deux processus distincts. Là où deux processus séparés doivent utiliser un mécanisme fourni par le système pour communiquer, les threads partagent une partie de l'état du processus, notamment sa mémoire. Dans le cas de données en lecture seule, il n'y a même pas besoin du moindre mécanisme de synchronisation pour que les threads utilisent les mêmes données. L'idée de base de multithreading est de permettre à plusieurs threads de partager les unités fonctionnelles d'un processeur unique de manière superposée. Le multithreading permet à un autre thread prêt de se remplir et évite ces ressources inactives, afin d'améliorer l'utilisation des ressources [18].

4.3. Inconvénients

La programmation utilisant des *threads* est toutefois plus rigoureuse que la programmation séquentielle, et l'accès à certaines ressources partagées doit être restreint par le programme lui-même, pour éviter que l'état d'un processus ne devienne temporairement incohérent, tandis qu'un autre thread va avoir besoin de consulter cette portion de l'état du processus. Il est donc obligatoire de mettre en place des mécanismes de synchronisation (à l'aide

de sémaphores, par exemple), tout en conservant à l'esprit que l'utilisation de la synchronisation peut aboutir à des situations d'interblocage si elle est mal utilisée.

La complexité des programmes utilisant des threads est aussi nettement plus grande que celle des programmes déferant séquentiellement le travail à faire à plusieurs processus plus simples (la complexité est similaire dans le cas de plusieurs processus travaillant en parallèle). Cette complexité accrue, lorsqu'elle est mal gérée lors de la phase de conception ou de mise en œuvre d'un programme[18].

4.4. Multithreading versus Multi-processing

Nombreux sont ceux qui ne font pas la différence entre multithreading et multiprocessing. Il s'agit donc d'une explication simplifiée pour ce point.

Le Multiprocessing fait référence à la capacité d'un système à prendre en charge plusieurs processeurs à la fois. Les applications d'un système multi-processing sont divisées en petites routines exécutées indépendamment. Le système d'exploitation alloue ces threads aux processeurs afin d'améliorer les performances du système[18].

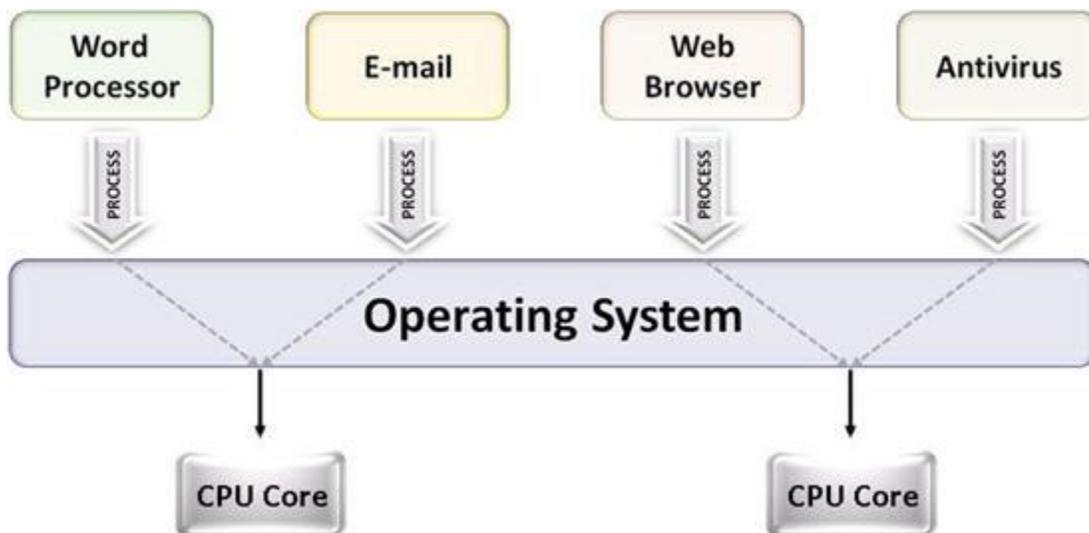


Figure 17 : multiprocessing en dual-core système

Un processeur est dit multithread s'il est capable d'exécuter efficacement plusieurs threads simultanément.

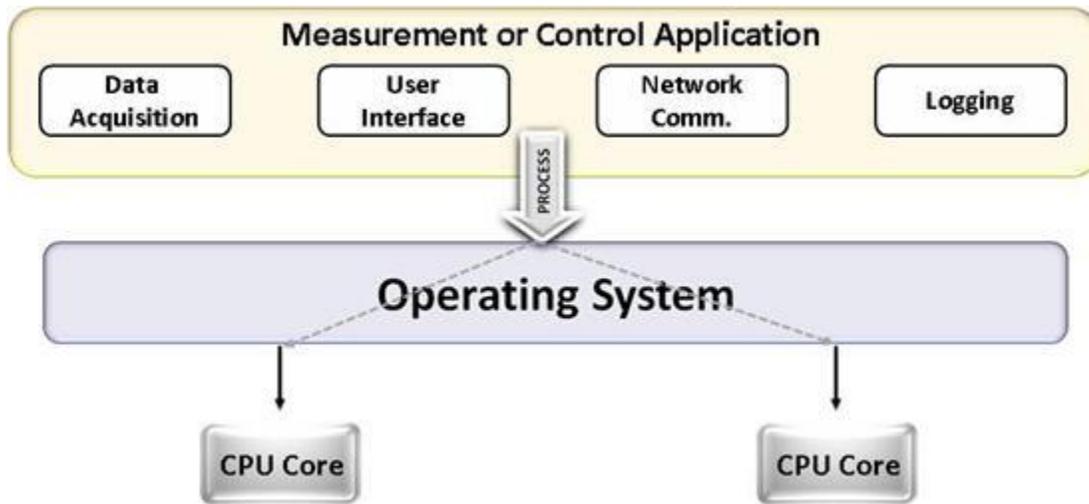


Figure 18 : multithreading en dual-core système.

Et le tableau suivant présente les différences entre eux :

Processeur	thread
Processus lourd ou gourmand en ressources	Le thread est léger et prend moins de ressources qu'un processus.
La commutation de processus nécessite une interaction avec le système d'exploitation	La commutation de threads n'a pas besoin d'interagir avec le système d'exploitation.
Dans plusieurs environnements de traitement, chaque processus exécute le même code mais dispose de ses propres ressources en mémoire et en fichiers.	Tous les threads peuvent partager le même ensemble de fichiers ouverts, processus enfants.
Si un processus est bloqué, aucun autre processus ne peut être exécuté tant que le premier n'est pas débloqué	Lorsqu'un thread est bloqué et en attente, le second thread de la même tâche peut être exécuté.
Plusieurs processus sans utiliser de thread utilisent plus de ressources.	Plusieurs processus threadés utilisent moins de ressources.
Dans de multiples processus, chaque processus fonctionne indépendamment des autres.	Un thread peut lire, écrire ou modifier les données d'un autre thread.

Tableau 3 : la différence entre multithreading et multiprocessing

5. Conclusion :

Au cours de ce chapitre, nous avons fourni une base théorique sur le Cloud Computing, en présentant ses types, ses services (IaaS, PaaS, SaaS), ses avantages et inconvénients, après discuté sur le multithreading qui aide dans le passage à l'échelle, mais le problème que se pose si les fournisseurs de cloud computing et paiement et orienté vers les entreprises.

Chapitre n°3 :

Expérimentation

1. Environnement de travail

1.1. Technologie utilisées

Nous présentons dans ce chapitre les technologies principales que nous avons utilisées pour réaliser notre travail d'expérimentation.

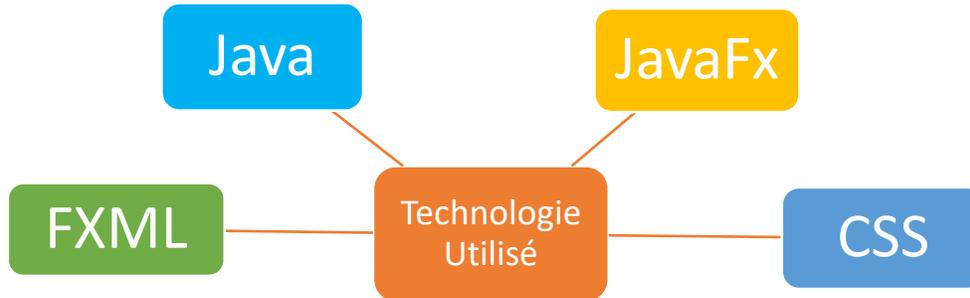


Figure 19 : technologie utiliser

1.1.1. Java

Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Beaucoup d'applications et de sites Web ne fonctionnent pas si Java n'est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable. Des ordinateurs portables aux centres de données, des consoles de jeux aux superordinateurs scientifiques, des téléphones portables à Internet, la technologie Java est présente sur tous les fronts [20].

1.1.2. JavaFx

Est une technologie créée par Sun Microsystems qui appartient désormais à Oracle, est un Framework Java permettant de construire des RIA (Rich Internet Application), JavaFX devient l'outil de création d'interface graphique GUI officiel du langage Java, pour toutes les sortes d'application (applications mobiles, applications sur poste de travail (desktop app), applications Web...).

JavaFX contient des outils très divers, notamment pour les médias audio et vidéo, le graphisme 2D et 3D. Le SDK de JavaFX étant désormais intégré au JDK standard Java SE, il n'y a pas besoin de réaliser d'installation spécifique pour JavaFX [20].

1.1.3. CSS

Le terme **CSS** est l'acronyme anglais de *Cascading Style Sheets* qui peut se traduire par feuilles de style en cascade. Le CSS est un langage informatique permettant de gérer la présentation et la mise en forme des applications créées avec javafx. (Positionnement des éléments, couleurs, tailles et polices, etc...).

1.1.4. FXML

FXML est un langage de balisage d'interface utilisateur basé sur XML créé par Oracle Corporation pour définir l'interface utilisateur d'une application JavaFX.

Il fournit une alternative pratique à la construction de tels graphiques en code procédural, et est idéalement adapté à la définition de l'interface utilisateur d'une application JavaFX, puisque la structure hiérarchique d'un document XML est très proche de la structure du graphe de scène JavaFX. Cependant, tout ce qui est créé ou implémenté dans FXML peut être exprimé directement à l'aide de JavaFX.[21]

1.2. Présentation des outils de développement

Nous présentons dans ce chapitre les outils principales que nous avons utilisées pour réaliser notre application.



Figure 20 : outils de développement

1.2.1. Netbeans

est un environnement de développement intégré (EDI) , un outil pour les programmeurs pour écrire, compiler, déboguer et déployer des programmes. Il est écrit en Java, ayant un succès et une base d'utilisateur très large, une communauté en croissance constante, et près 100 partenaires mondiaux et des centaines de milliers d'utilisateur à travers le monde. Sun Microsystems a fondé le projet open source NetBeans en Juin 2000 et continue d'être le sponsor principal du projet [21].

1.2.2. SceneBuilder

Le JavaFX SceneBuilder est un outil initialement créé par Oracle et désormais faisant partie de l'OpenJFX qui permet de créer des fichiers au format FXML via un éditeur graphique. Cet outil est disponible en tant qu'application autosuffisante qui peut être lancée depuis votre bureau ou en tant qu'API intégrable dans des outils tiers tels que NetBeans [21].

1.3. Bibliothèque utilisée

Pour développer l'application la figure ci-dessous montre l'ensemble des bibliothèques utilisées :

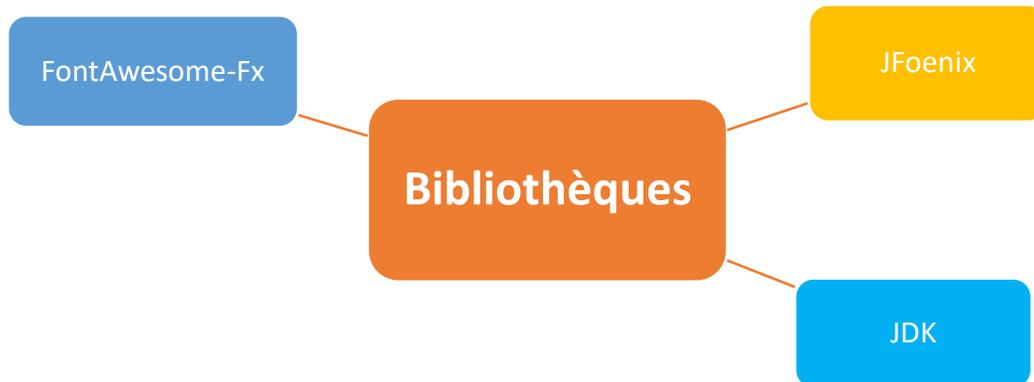


Figure 21 : bibliothèque utilisée

1.3.1. JDK

Le Java Développement Kit (JDK) désigne un ensemble de bibliothèques logicielles de base du langage de programmation Java, ainsi que les outils avec lesquels le code Java peut être compilé [22].

1.3.2. FontAwesome-Fx

Cette bibliothèque fournit des icônes, et prend également en charge le style CSS qui nous aide à personnaliser les icônes de façon personnalisable .

1.3.3. JFoenix

JFoenix est une bibliothèque Java open source, qui implémente Google Matériel Design tout en utilisant des composants Java [22].

2. jeu de données

Afin de pouvoir appliquer notre démarche, nous avons besoin d'un jeu de données contenant les utilisateurs, les items à recommander et les ratings donnés pour ces items. Pour cela nous allons utiliser la base MovieLens2 (ml-latest-small) qui a été collectée par

le projet de recherche GroupLens à l'Université de Minnesota. Cette base contient les informations de 200000 films, 610 utilisateurs.

Le fichier rating.csv contient des votes créés par des utilisateurs sur des films (items). C'est une liste dont les champs suivants :

Champs	Description
userId	L'identifiant de chaque utilisateur.
movieId	L'identifiant du film
Rating	La note donnée par un utilisateur à un film.

Tableau 4 : Tableau de description des champs du fichier rating.csv

3. Mesures de tests

Nous avons utilisé la mesure MAE (Mean Absolute Error) dans notre évaluation, qui est également une mesure d'erreur permettant de calculer la moyenne des erreurs absolues entre les prédictions et les valeurs réelles fournies par les utilisateurs. En d'autres termes, elle permet de déterminer la qualité de ces prédictions. Rappelons que les prédictions sont générées à partir d'un ensemble d'apprentissage. De plus, cette mesure est largement utilisée par les chercheurs afin de tester leurs algorithmes dans le domaine du filtrage d'information. Nos expérimentations sont effectuées sur plusieurs configurations (tableau 5).

Configuration	Type de processeur	Fréquences (GHZ)	Capacité de RAM (GB)
Config 1	i3	1.7	4
Config 2	i5	1.7	8
Config 3	i7	2.4	8
Config 4	I7	2.4	16

Tableau 5 : configuration matérielles

4. Résultats

Dans cette section, nous allons présenter différentes captures issues de notre application. Elle est dotée d'une interface graphique facile, pour suivre la démarche adoptée, et voir la sérialisation de réalisation.

La figure ci-dessous présente la fenêtre principale de notre application.

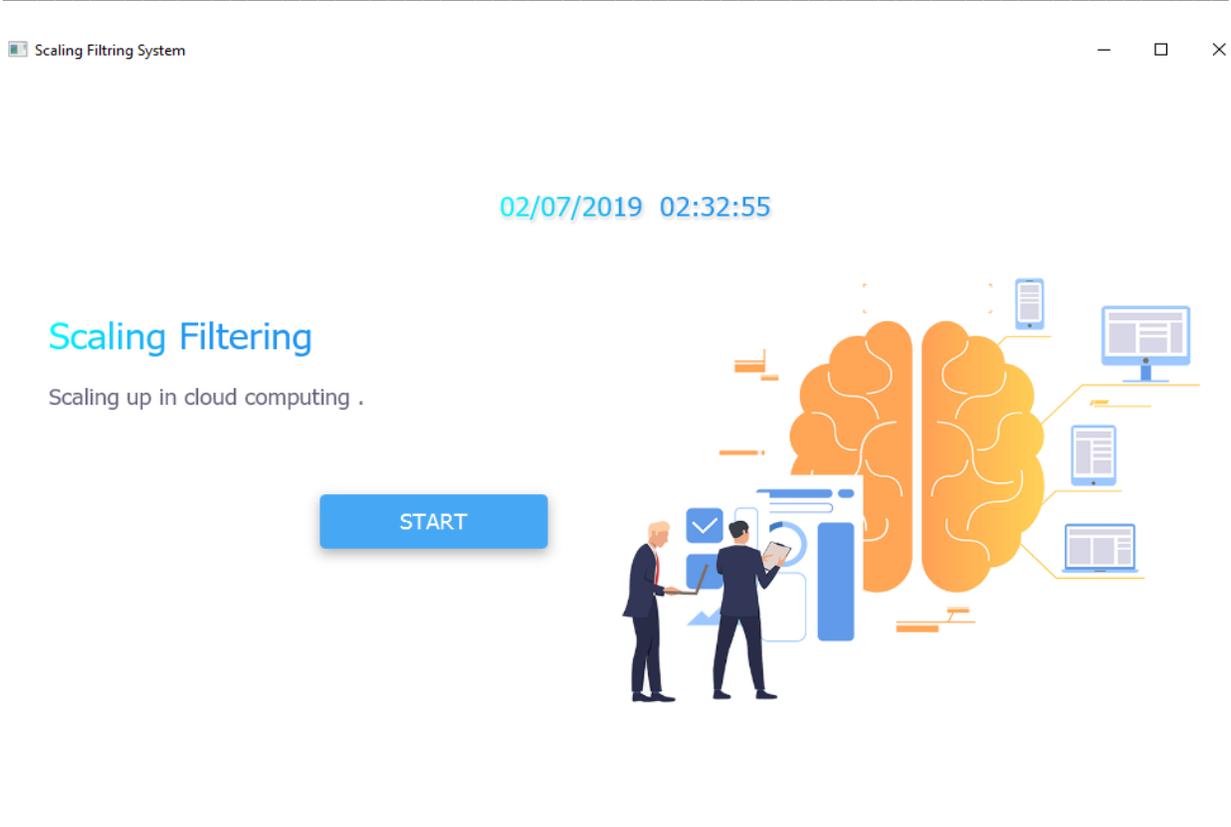


Figure 22 : page d'accueil

Alors on appuie sur le botton start pour accéder à la fenêtre de chargement de données.

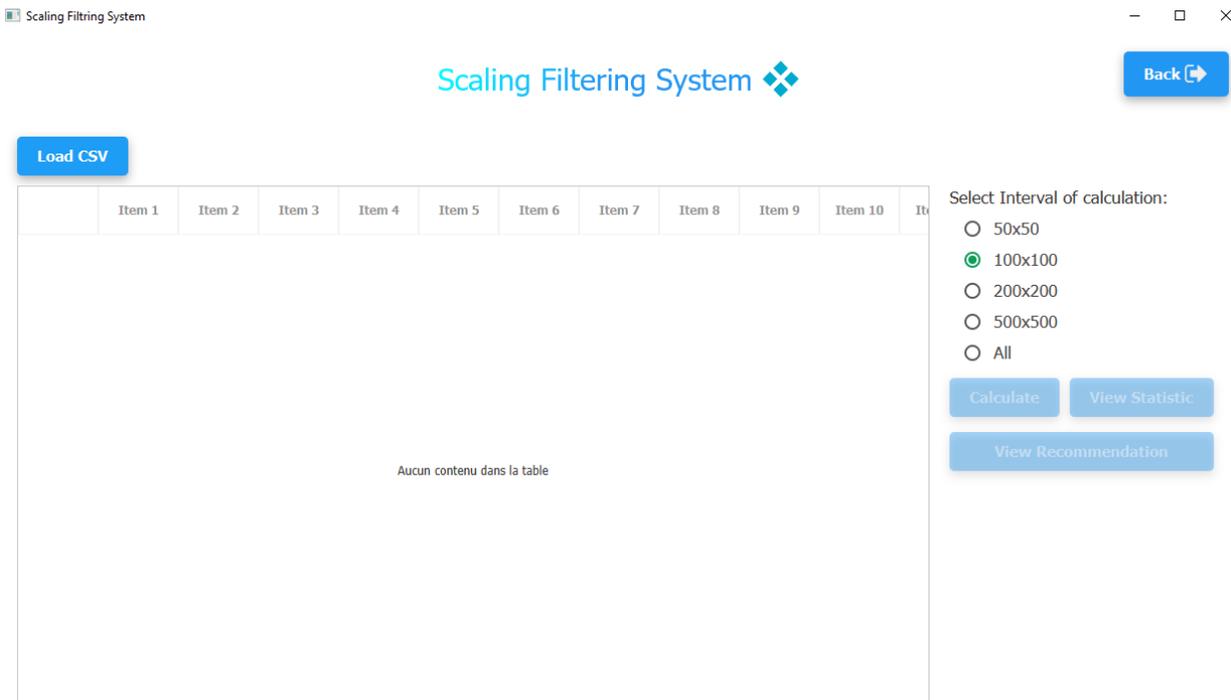


Figure 23 : page chargement de données

Par la suite nous devons appuyer le bouton **load csv** pour ouvrir les emplacements des fichiers tests.

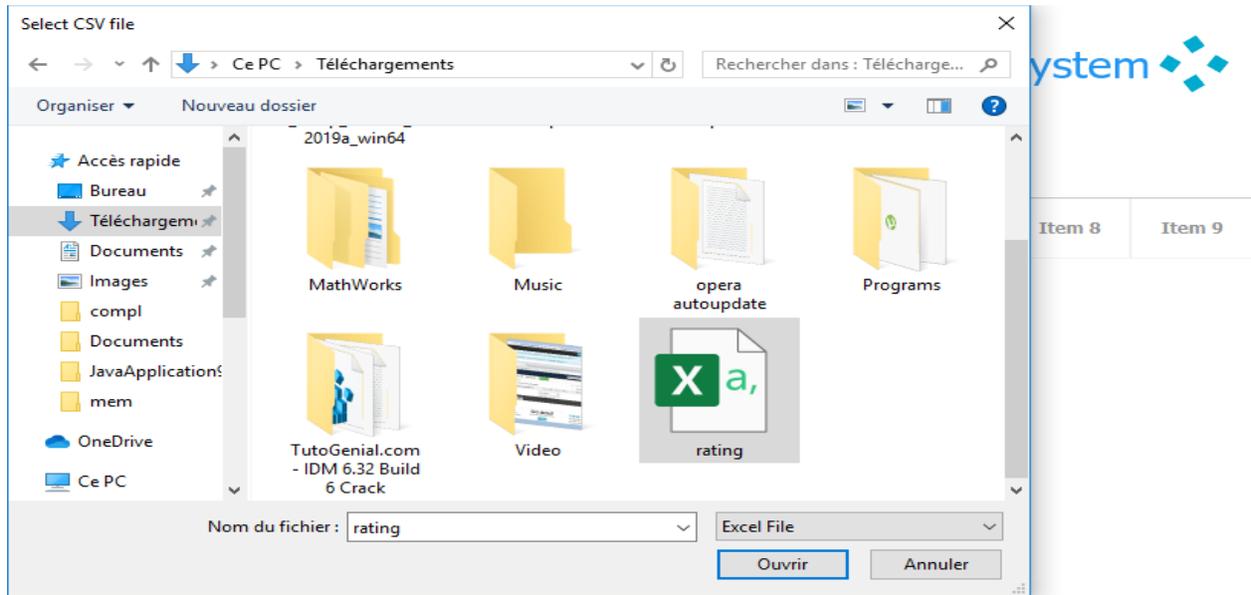


Figure 24 : choix du fichier rating.csv

Une fois le fichier choisi une matrice d'évaluation sera affichée (figure 25)

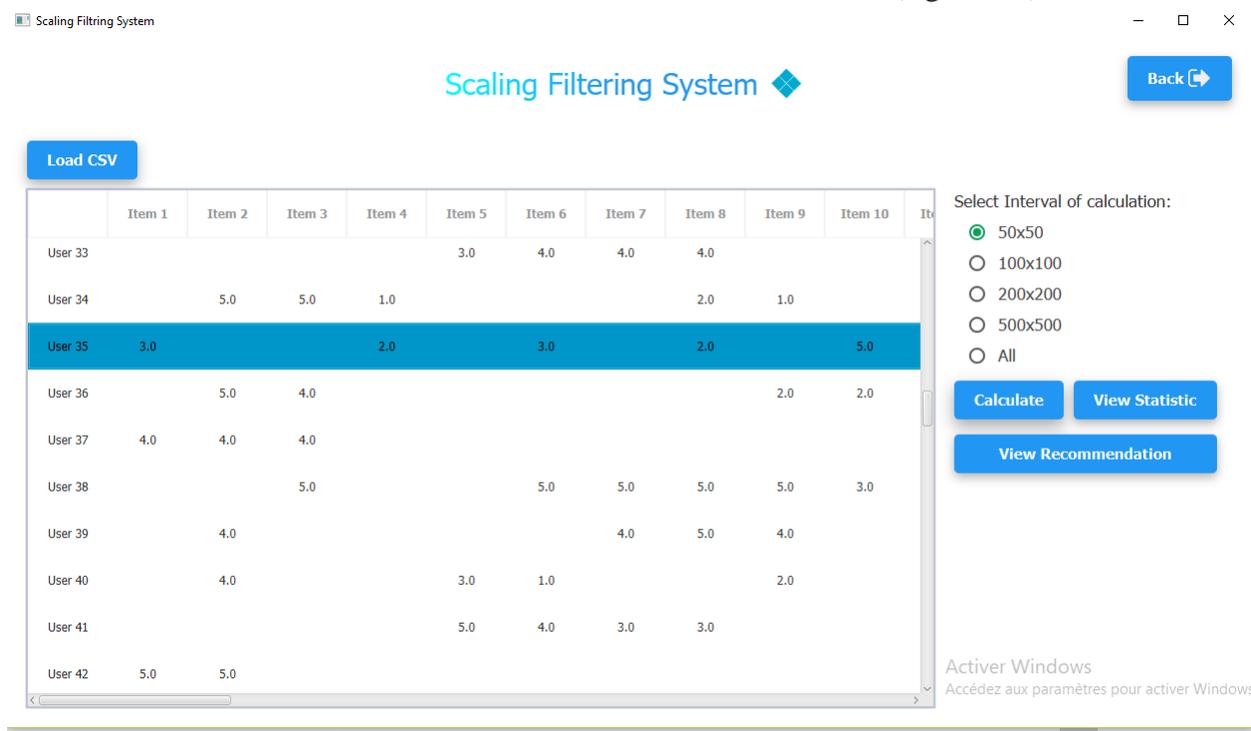


Figure 25 : matrice d'évaluation

On remarque que le bouton calculate devenir activé à l'utilisation, avant de déclencher le processus de calcul il faut choisir la taille de la matrice à mettre en ouvre.

Select Interval of calculation:

- 50x50
- 100x100
- 200x200
- 500x500
- All

Calculate

View Statistic

View Recommendation

Figure 26 : Choix de la taille de la matrice

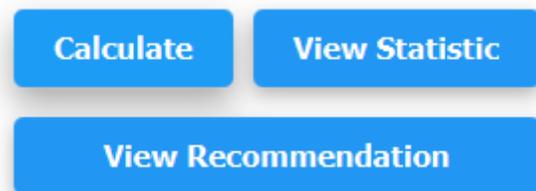
Puis on procédé le calcul (*calculate*)



Figure 27 : l'interface du résultat

Dans cette interface on affiche la table de similarité, prédiction et le temp d'exécution.

Après le calcul, il faut acheminer l'ensemble des étapes via boutons **view static** et **view recommendation** et l'interfacemontrant l'erreur est affichée.



Statistics

Average Error

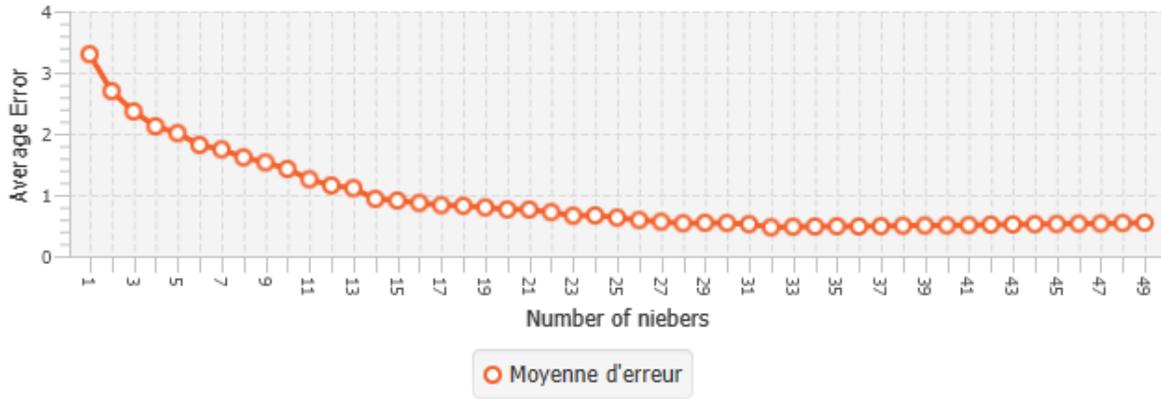


Figure 28 : MAE

Hardware speed comparaison

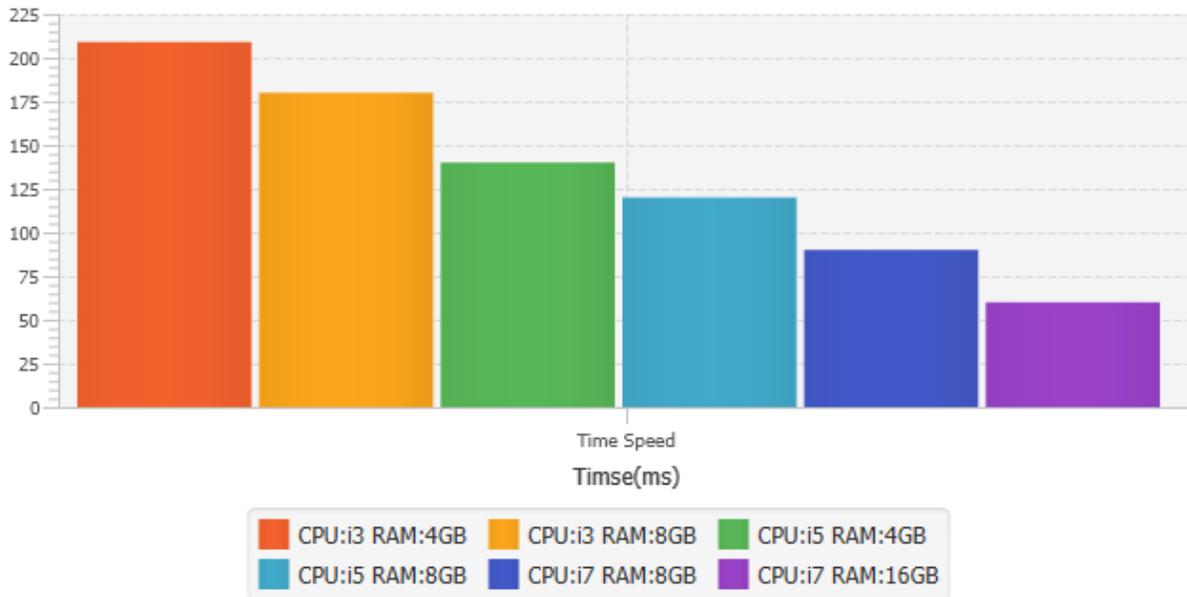


Figure 29 : comparaisons entre des déférenttes configurations

Enfin on click sur le bouton **view recommendation** et l'interface ci-dessous est affichée

Recommendation ✕

Id user	Id of item recommended																								
User 1	6	9	11	12	13	14	15	16	20	23	27	31	39	42	44	47	49	53	54	56	60	61	62	64	69
User 2	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22	23	26	27	28	29	30	31	32	33
User 3	6	10	12	14	17	19	20	21	22	23	25	26	27	28	29	31	32	37	38	39	40	42	45	47	48
User 4	3	6	10	12	14	19	20	22	23	25	26	27	28	29	31	32	35	37	38	39	40	41	42	45	47
User 5	2	6	8	9	10	11	12	14	15	16	18	22	23	27	30	31	33	35	37	42	45	48	49	51	52
User 6	2	4	5	10	11	14	15	16	17	18	19	20	23	25	27	28	29	30	31	32	35	37	38	39	40
User 7	3	5	7	8	10	13	14	15	16	18	19	21	22	24	26	28	30	32	33	35	36	40	41	42	48
User 8	2	6	10	14	15	16	19	21	22	23	26	27	28	31	32	39	40	44	45	48	49	50	52	56	57
User 9	2	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	27	28	30	31
User 10	1	2	3	4	5	6	7	8	10	13	14	15	16	18	19	20	21	22	24	25	26	27	28	29	30

Figure 30 : l'interface de recommandation

Pour avoir l'ensemble des items recommandés à un utilisateur donné par son identifiant la liste des recommandations est affichée dans l'onglet (figure 31)



Figure 31 : Liste de recommandation

Faute de mode de paiement fortement exigé par les différents fournisseurs du cloud ce que nous a empêché de bénéficier cette opportunité et s'opposer de migrer notre expérimentation en cloud computing.

5. Conclusion

Dans ce dernier chapitre, nous avons exposé la réalisation de notre application qui tente d'avoir l'impact du passage à l'échelle sur le temps d'exécution de l'algorithme.

Nous avons itéré l'algorithme basé mémoire en mode (item-item) par différentes tailles de matrices et selon 04 configuration distinctes pour avoir l'effet de passage à l'échelle qui est un problème pertinent dans les systèmes de recommandation. Nous avons constaté que ce type d'algorithme est gourmand en ressources ce qui demande des machines haute gamme pour satisfaire les utilisateurs du web.

Conclusion générale

Un système de filtrage collaboratif (SCF) permet la découverte des items intéressants, grâce à l'authentification du processus de la recommandation. Parmi les avantages offerts aux utilisateurs, on trouve :

- La possibilité d'exprimer leur avis quant à la pertinence des items, selon leurs goûts et la qualité qu'ils perçoivent sur ces items ;
- La possibilité de recevoir des recommandations inattendues car il suffit qu'un utilisateur, de profil proche, les ait jugées intéressantes ;
- La possibilité de bénéficier des évaluations sur les items que d'autres utilisateurs membres de sa communauté ont déjà faites.

Tous ces avantages sont apportés aux utilisateurs par le principe de collaboration, en contrepartie d'un effort individuel. Ces systèmes requièrent la participation active des utilisateurs, sur le long terme, pour atteindre de bonnes performances.

Néanmoins, ces systèmes sont limités par certaines lacunes comme le problème de passage à l'échelle due à la complexité des calculs. Ce problème est typiquement due au coût élevé affecté par le comptage des voisins notamment pour les applications de e-commerce où le nombre des internautes est très élevé. Or, le système doit fournir en fraction de seconde des recommandations prévues pertinentes à l'utilisateur actif.

Notre objectif était d'avoir l'impact de telles ressources matérielles afin d'optimiser le temps de calcul. De plus une solution logicielle a été proposée c'est le multi-threading, pour que chaque noyau du processeur s'occupe d'une tâche séparément.

Actuellement, la solution cloud se révèle la plus fructueuse compte tenu des moyens que s'y possèdent. L'inconvénient majeur de cette solution est qu'elle est purement lucrative et destinée aux grandes entreprises.

Bibliographes

- [1] Herlocker, J. L., Konstan, J. A., and Riedl, J., Explaining collaborative filtering recommendations. In Proceedings of the 2000 Conference on Computer Supported Cooperative Work, 2000.
- [2] MALTZ D., EHRLICH K., « Pointing the way: active collaborative filtering », Proceedings of CHI'95, p. 7-11, mai 1995.
- [4] Adomavicius, G. Tuzhilin, A., Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng., 2005.
- [8] : Landry Fossou Noumsi, étude et mise en place d'une solution Cloud Computing, école national supérieur des postes et des télécommunications, 2012.
- [9] : V. Kherbache, M. Moussalih, Y. Kuhn, A. Lefort, Cloud Computing, IUT Nancy Charlemagne.2009/2010.
- [10] : N.Grevet. Le cloud computing : évolution ou révolution ? Pourquoi, quand, comment et surtout faut-il prendre le risque ?
- [11] : J.Zhu.Cloud Computing Technologies and Applications. Springer Science+Business Media, LLC 2010.
- [13] : Le Cloud Computing, Définition et impact pour les SSII, 2012.
- [15] : Civic Consulting. L'informatique en nuage, étude Département thématique a : Politiques Economiques et scientifiques. 2012.
- [16] : A El Mahdaouy and M Oumsis. Evaluation et amélioration de performances des algorithmes d'équilibrage de charges dans un environnement cloud computing. casablanca, morocco, 2013.
- [17] : Mohand Boughanem, LyndaTamine-Lechani, JoséMartinez, SylvieCalabretto, JenaPierreChevallet. Un nouveau passage à l'échelle en recherche d'information. Revue des Sciences et Technologies de l'Information - Série ISI: Ingénierie des Systèmes d'Information, Lavoisier, 2006.

Webographie

- [3] Mathieu, Les algorithmes de recommandation, <http://www.podcastscience.fm/dossiers/2012/04/25/les-algorithmes-de-recommandation/>
- [5] <http://mrim.imag.fr/publications/2003/CB001/berrut03b.pdf>
- [6] : <http://www.partagedefichier.com/blog/cloud-computing-definition/>. Juin 2012.
- [7] : <https://blog.blaisethirard.com/qu-est-ce-que-le-cloud-computing/>
- [12] : <https://csrc.nist.gov/Projects/Cloud-Computing>
- [14] : https://d1.awsstatic.com/International/fr_FR/whitepapers/the-path-to-the-cloud-dec2015.pdf
- [18] <https://fr.wikipedia.org/wiki/Multithreading>
- [19] <https://www.newgenapps.com/blog/top-5-cloud-platforms-and-solutions-to-choose-from>
- [20] <https://www.java.com/fr>
- [21] <https://docs.oracle.com>
- [22] <https://fr.wikipedia.org>