

Étude comparative de Top_k basée sur l'algorithme de Fagin en utilisant des métriques de corrélation dans la qualité de service de Cloud Computing

Kaoutar El Handri*, Abdellah Idrissi**

*, ** Département Informatique,
Équipe IPSS, Faculté de science de Rabat, Université Mohammed V, Maroc
* kaoutar.elhandri@um5s.net.ma
** abdellah.idrissi@um5.ac.ma

1 Introduction

Aujourd'hui, il est crucial d'extraire seulement les données les plus recommandés d'informations. Plusieurs travaux se sont focalisés sur l'extraction de k-dominance comme l'étude des algorithmes Top_k . (Amagata et al., 2018). Dans ce contexte, l'approche représenté dans le présent papier se base sur un algorithme Top_k qui affine le résultat de Skyline (Idrissi et al., 2016). La mise en place de cette optimisation est alors basée sur l'interprétation des préférences des utilisateurs du Cloud Computing (CC), en utilisant des méthodes d'Analyse Multi-critères (ADMC), particulièrement la Méthode de la Somme Pondéré (MSP) tout en donnant naissance à l'algorithme ($Top_{k_{WS}}$). Ensuite, afin d'évaluer la performances de notre approche, nous avons comparé l'algorithme proposé avec celui de Fagin (FA) (El handri et Idrissi, 2018).

2 L'approche proposée et les résultats d'expérimentation

L'objectif de $Top_{k_{WS}}$ se base sur le calcul des services Cloud les plus pertinents basé sur la QoS. L'adaptatif MSP dans la fonction 1 a pour but de refléter correctement la préférence de l'utilisateur final avec l'agrégation de deux fonctions monotones et linéaires comme une seule fonction score.

$$f : \max \sum (w_j a_{ij}) + \max \sum (w_j (1/a_{ij})) \quad (1)$$

Tel que les poids $\omega_i, i \in \{1, \dots, p\}$ et $\sum_{i=1}^p \omega_i = 1, \omega_i \geq 0, i \in \{1, \dots, p\}$, sachant que les a_{ij} sont les critères de la QoS de service Cloud Computing. Ensuite, nous avons étudié l'impact des différentes tailles de données d'entrée ainsi que l'impact de nombre de critères choisis sur le temps d'exécution. La figure 1 montre que l'algorithme $Top_{k_{WS}}$ a une meilleure performance que l'algorithme de Fagin. De plus, l'étude de corrélation résumé dans la table 1 prouve cette performance (El handri et Idrissi, 2018), en se basant sur les métriques de recommandation (CCP), (CCS), et (CCK) qui sont respectivement le Coefficient de Corrélation de Pearson, de Spearman, et de Kendall. Ensuite pour valider ces résultats nous avons appliqué le test de significativité basé sur la p-value et le test de Student (El handri et Idrissi, 2018).

Étude comparative de Top_k basée sur l'algorithme de Fagin

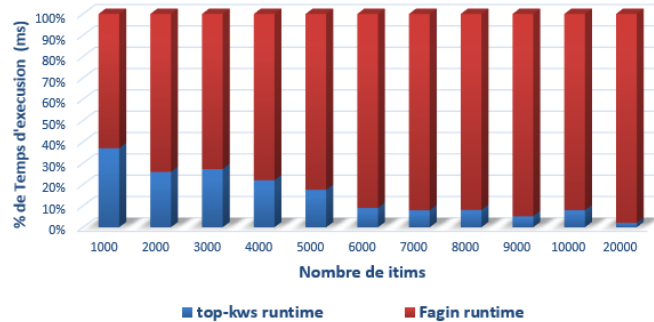


FIG. 1 – Comparaison de pourcentage de temps d'exécution de l'algorithme Top_{kWS} avec celui de FA en fonction de la taille des entrées

	CCP	CCS	CCK
Top_{kWS}	72%	36%	46%
FA	28%	64%	54%

TAB. 1 – Comparaison entre Top_{kWS} et FA en se basant sur les trois métriques de corrélation.

3 Conclusion

Dans cet article, nous avons présenté notre approche basée sur l'algorithme Top_{kWS} pour affiner le Skyline. Nous soulignons l'amélioration significative de temps d'exécution donné par notre algorithme comparé à l'algorithme de Fagin, en essayant de trouver un compromis entre un bon temps de réponse et une bonne qualité des résultats trouvés. Comme perspectives, Nous compterons à paralléliser cet algorithme dans un contexte distribué basé sur la technologie de Big Data comme Hadoop et Spark.

Références

- Amagata, D., T. Hara, et M. Onizuka (2018). Space filling approach for distributed processing of top-k dominating queries. *IEEE Transactions on Knowledge and Data Engineering* 30(6), 1150–1163.
- El handri, K. et A. Idrissi (2018). Comparative study of topk based on fagins algorithm using correlation metrics in cloud computing qos, 2018. *International Journal of Internet Technology and Secured Transactions (IJITST)*, (in press).
- Idrissi, A., K. Elhandri, H. Rehioui, et M. Abourezq (2016). Top-k and skyline for cloud services research and selection system. In *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, pp. 40. ACM.