

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسبدي بلعباس  
Ecole Supérieure en Informatique  
-08 Mai 1945- Sidi Bel Abbas



## MÉMOIRE

En vue de l'obtention du diplôme de **Master**

Filière : **Informatique**

Spécialité : **Systeme d'Informations et Web (SIW)**

## Thème

---

Etude des techniques d'estimation de cardinalité pour  
l'optimisation des requêtes dans un contexte relationnel

---

Présenté par :

- Mlle BOUROKBA Serine
- Mme MERZOUG Fatima

Soutenu le : **26/09/2021**

Devant le jury composé de :

- |                       |            |           |
|-----------------------|------------|-----------|
| - Mr KESKES Nabil     | Professeur | Président |
| - Mr KECHAR Mohamed   | Docteur    | Encadreur |
| - Mr KAZI TANI Yacine | Docteur    | Examineur |

*Année Universitaire : 2020 / 2021*

## Résumé

L'épine dorsale des entreprises de nos jours, qui sont caractérisées par un grand volume d'information est les systèmes de gestion de bases de données relationnelles (SGBDR). Ces systèmes fournissent un environnement confortable pour gérer les données tout au long de leur cycle de vie: la génération, le stockage, la maintenance et le traitement. Le SQL (Structured Query Language) est l'interface standard pour interroger Les données stockées dans les SGBDR.

Connaître la taille du résultat d'une requête SQL (c-à-d la taille en page disque ou en nombre de tuples), avant son exécution, peut aider à la fois le SGBDR à trouver le plan optimal nécessitant le moins de ressources système, ainsi que les utilisateurs en leur fournissant le résultat final en un minimum de temps. Cependant, cette tâche fondamentale qui est connu sous le nom de l'estimation de la cardinalité est considéré comme un problème difficile. Dans le contexte de notre notre travail, nous définissons l'estimation de la cardinalité comme l'estimation de la taille du résultat d'une requête SQL donnée.

Malheureusement, les approches basées sur le synopsis et l'échantillonnage pour l'estimation de la cardinalité qui sont largement utilisées dans les moteurs de bases de données industrielles donne parfois des erreurs catastrophiques dans la pratique. Cet écart a motivé un grand nombre de nouvelles études préconisant l'utilisation d'approches basées sur l'apprentissage automatique/l'apprentissage en profondeur pour l'estimation de la cardinalité. Ces nouvelles approches ont cependant leurs propres limitations, y compris un effort de formation important, l'incapacité à gérer les modifications de données dynamiques et la généralisation à des requêtes inconnues.

Dans ce travail, nous présentons une étude de quelques approches d'estimation de cardinalité (dans un contexte relationnel) qui ont été proposées dans la littérature. Nous présentons également une comparaison entre ces travaux et une synthèse pour extraire suffisamment d'informations dans ce domaine de recherche.

**Mot-clé :** SGBDR, Optimisation des requêtes, Estimation de la cardinalité.

**Abstract**

The backbone of today's information-rich society is Relational Database Management Systems (RDBMS), which provide a comfortable environment for managing enterprise data across its entire life cycle of generation, storage, maintenance, and processing. This data stored in the RDBMS is queried using the standard interface "Structured Query Language (SQL)". Prior to execution, knowing the predicted size of the SQL query result, as measured by output row-cardinality, can help both the RDBMS system and the user in a variety of ways like assessing query feasibility, approximating query answers, query progress monitoring, and resource allocation techniques. In the context of our work, we define cardinality estimation as the estimation of the result size (number of rows in the output) of a given SQL query.

Unfortunately, the synopsis and sampling-based approaches for cardinality estimate widely employed in industrial database engines are sometimes catastrophically wrong in practice. This gap has prompted a flurry of new studies arguing for the use of machine learning/deep learning approaches for cardinality estimation. These new approaches, however, have their own set of drawbacks, including important training effort, inability to handle dynamic data changes, and generalization to unknown queries.

In this work, we present a study of some cardinality estimation approaches (in a relational context) which have been proposed in the literature. We also present a comparison between this work and a synthesis to extract sufficient information in this area of research.

**Keywords :** RDBMS, Query optimization, Cardinality estimation.