

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 : **Systemes d'information et Web (SIW)**

Thème

Big Data Warehouse Optimization : Optimisation des entrepôts de données dans un environnement Big Data

Présenté par :

- Mr SLIMANI Ibrahim
- Mr BELABBES Mohamed Abdelillah

Soutenu le : **24/06/2019**

Devant le jury composé de :

- | | |
|-----------------------|-----------|
| - M. MALKI Mimoun | Président |
| - M. MALKI Abdelhamid | Encadreur |
| - M. KECHAR Mohamed | Examineur |

Année Universitaire : 2018 / 2019

Résumé

Les entrepôts de données constituent l'espace de stockage centralisé pour les systèmes décisionnels.

L'accès aux données entreposés dans cette structure est absolument complexe vue que les requêtes analytiques impliquent plusieurs opérations de jointures et de fonctions d'agrégation ce qui rend le temps de réponse très lent.

Pour cette raison il faut mettre en place des stratégies d'optimisation pour lever les performances et accélérer le temps d'exécution de requêtes à savoir les indexes, les vues matérialisés, la fragmentation...

Avec l'apparition de l'ère du Big Data les entrepôts de données de prochaine génération seront déployé sur une infrastructure et des architectures hétérogènes intégrant à la fois des données structurées traditionnelles et des données volumineuses non structurés.

Pour cette raison il faut s'adresser au techniques d'optimisation dans les bases NoSQL, et dans les solutions Big Data tels que Hadoop, apache Spark...

Mots clés: Data warehouse, Big Data, Optimisation, Hadoop, MapReduce, NoSQL

Abstract

Data warehouses are the centralized storage space for business intelligence systems.

Access to the data stored in this structure is absolutely complex as the analytic requests involve several join operations and aggregation functions which makes the response time very slow.

For this reason it is necessary to implement optimization strategies to lift the performance and speed up the execution time of queries namely indexes, materialized views, fragmentation ...

With the advent of the Big Data era, next-generation data warehouses will be deployed on heterogeneous infrastructure and architectures that integrate both traditional structured data and large, unstructured data.

For this reason we must address the optimization techniques in NoSQL databases, and in Big Data solutions such as Hadoop, Apache Spark ...

Keywords: Data warehouse, Big Data, Optimization, Hadoop, MapReduce, NoSQL