

Ecole Doctorale
Mathématiques et Sciences et Technologies de l'Information et de la Communication

*Laboratoire des Sciences et Techniques de l'Information, de la Communication
et de la Connaissance*

AVIS DE SOUTENANCE DE THESE

Le jeudi 20 décembre 2018 à 13h30

à l'UFR Sciences et Techniques, amphithéâtre "F", 6 avenue Victor Le Gorgeu, Brest

Monsieur LAGA AREZKI

soutiendra une thèse de doctorat sur le sujet suivant :

" Optimisation des performance des logiciels de traitement de données sur les périphériques de stockage
SSD ".

Le jury sera ainsi composé :

- **M. ACQUAVIVA JEAN-THOMAS, Docteur**
Data Direct Network (DDN) - MEUDON
- **M. BARAIS OLIVIER, Professeur des universités**
IRISA - INRIA - RENNES
- **M. BOUGANIM LUC, Directeur de Recherche**
INRIA Saclay et UVSQ - VERSAILLES
- **M. BOUKHALFA KAMEL, Professeur**
USTHB - HYDRA - ALGERIE
- **M. BOUKHOBZA JALIL, Maître de conférences**
Univ. de Bretagne Occidentale - BREST
- **MME SAHRI SOROR, Maître de conférences**
Université Paris Dauphine - PARIS 16EME
- **M. SINGHOFF FRANK, Professeur des universités**
Univ. de Bretagne Occidentale - BREST

A BREST, le 26 novembre 2018

Le Président de l'Université
de Bretagne Occidentale,



M. GALLOU

Optimisation des performances des logiciels de traitement de données sur les périphériques de stockage SSD

Mots clés : base de données, tri, pré-chargement de données, mémoire flash, SSD ...

Résumé :

"Nous assistons aujourd'hui à une croissance vertigineuse des volumes de données. Cela exerce une pression sur les infrastructures de stockage et les logiciels de traitement de données comme les Systèmes de Gestion de Base de Données (SGBD).

De nouvelles technologies ont vu le jour et permettent de réduire la pression exercée par les grandes masses de données. Nous nous intéressons particulièrement aux nouvelles technologies de mémoires secondaires comme les supports de stockage SSD (Solid State Drive) à base de mémoire Flash.

Les supports de stockage SSD offrent des performances jusqu'à 10 fois plus élevées que les supports de stockage magnétiques. Cependant, ces nouveaux supports de stockage offrent un nouveau modèle de performance. Cela implique l'optimisation des coûts d'E/S pour les algorithmes de traitement et de gestion des données.

Dans cette thèse, nous proposons un modèle de calcul des coûts d'E/S sur SSD pour les algorithmes de traitement de données. Ce modèle considère principalement le volume des données, l'espace mémoire alloué et la distribution des données.

Nous proposons également un nouvel algorithme de tri en mémoire secondaire : MONTRES. Ce dernier est optimisé pour réduire le coût des E/S lorsque le volume de données à trier fait plusieurs fois la taille de la mémoire principale.

Nous proposons enfin un mécanisme de pré-chargement de données : Lynx. Ce dernier utilise un mécanisme d'apprentissage pour prédire et anticiper les prochaines lectures en mémoire secondaire."

Title: Performance optimization for Data processing softwares on SSD storage devices

Keywords: Database, sorting, prefetching, flash memory, SSD

abstract:

"The growing volume of data poses a real challenge to data processing software like DBMS (DataBase Management Systems) and data storage infrastructure.

New technologies have emerged in order to face the data volume challenges. We considered in this thesis the emerging new external memories like flash memory-based storage devices named SSD (Solid State Drive).

SSD storage devices offer a performance gain compared to the traditional magnetic devices. However, SSD devices offer a new performance model that involves IO cost optimization for data processing and management algorithms.

We proposed in this thesis an IO cost model to evaluate the data processing algorithms. This model considers mainly the SSD IO performance and the data distribution.

We also proposed a new external sorting algorithm: MONTRES. This algorithm includes optimizations to reduce the IO cost when the volume of data is greater than the allocated memory space by an order of magnitude.

We proposed finally a data prefetching mechanism: Lynx. This one makes use of a machine learning technique to predict and to anticipate future access to the external memory."