Titre : Stratégies efficaces d'accès aux données sur bandes magnétiques.
Title: Effective Strategies for Accessing Data on Magnetic Tapes.

Mots-clés : Analyse de performance, stockage, bandes magnétique
Keywords: Performance analysis, storage, magnetic tapes

Profil et compétences recherchées : Connaissances en analyse de performance et programmation système bas-niveau, avec une bonne connaissance de l'écosystème HPC.
Required skills : Prior knowledge on low-level programming performance analysis, with a strong background of the HPC ecosystem.

Description détaillée du sujet en français

Les bandes magnétiques sont un support de stockage indispensables à la gestion les immenses volumes de données. Leur utilisation couvre de nombreux domaines, des institutions scientifiques de premier plan (CERN [1], CEA, etc.) à l'industrie (fournisseurs Cloud et *hyperscalers*, prévisions météorologiques [2], etc.). L'importance des bandes magnétiques vient de leurs nombreux avantages opérationnels : durée de vie, consommation d'énergie, rapport coût/octet, etc. Ces nombreux avantages sont contrebalancés par un temps d'accès aux données bien supérieurs à celui sur disques (de l'ordre de la minute pour les bandes). Peu de recherches académiques ont été menées sur la mesure de performance et l'optimisation des coûts opérationnels du stockage sur bande, contrairement aux disques durs mécaniques et SSD qui bénéficient depuis plusieurs décennies d'études approfondies. L'objectif de cette thèse est d'étudier en profondeur les processus physiques et les interactions logicielles d'un système stockage sur bandes, et de proposer de nouvelles approches visant à optimiser les temps de lecture des données sur bande.

PhD thesis subject in English

Magnetic tapes are an essential storage medium for managing massive amounts of data. They are used in a wide range of fields, from leading scientific institutions (CERN [1], CEA, etc.) to industry (cloud providers and *hyperscalers*, weather forecasting [2], etc.). The importance of magnetic tape stems from its many operational advantages: lifespan, power consumption, cost per byte, etc. These numerous advantages are offset by data access times that are significantly longer than those of conventional disks (on the order of a minute for tapes). Little academic research has been conducted on performance measurement and the optimization of operational costs for magnetic tape storage, unlike mechanical hard drives and SSDs, which have been the subject of in-depth studies for several decades. The objective of this thesis is to study, at a fine-grained level, the physical processes and software interactions involved in magnetic tape operations, and to propose new approaches for optimizing data recall times on tape.

Context / Contexte

Magnetic tapes have been used to store computer data since the 1950s, so the layman now often considers it as an outdated technology. However, tape storage is still and will remain essential in various domains, from academic institutions to industrial actors. For instance, CERN and CEA are leading scientific institutions that store exabytes of data, hence heavily rely on tapes due to a limited budget.

Tape technology remains crucial in the storage stack due to the many advantages it offers. They do not consume energy when they are not used, offer a strong security (offline storage in robotic libraries), and exhibit a better longevity for a much lower cost per TB than disks. Tape capacity is expected to witness a massive leap in coming years, increasing to several hundreds of TB per tape. This evolution will amplify the main benefits of tape storage.

However, tapes have often been primarily considered for archiving cold data, because of their main drawback: it takes around a minute to mount a tape from its shelf into a drive and position the reading head before starting reading data. This explains the current lack of academic effort to optimize relatively frequent data access. Nevertheless, more and more research projects require handling tremendous volumes of data, which are not only destined to be archived but also regularly accessed for scientific processing.

Objectives / Objectifs

The general idea of the project is to increase IO throughput at the tape level, so that the performance of the global Hierarchical Storage Management system (HSM) is improved. The methodology is to start from the tape level itself, and leverage knowledge on its performance to enhance the overall IO thoughput.

In this context, the first objective of the thesis is to evaluate the performance of tape operations. We plan to understand the low-level interaction between the tape drive, the tape cartridge and the connected device. The approach that we propose in this thesis is to measure the operational cost of tape operations at a fine-grained level. One of the considered approaches will be to introduce performance counters and metrics inside the tape firmware and the related software controlling the tape drive. This could be done by taking an approach similar to implementing SMART counters [3] in HDDs.

The second objective is to propose new APIs for more efficient operations, building on the results of the first objective. Those APIs will leverage knowledge of the mechanical processes involved in tape operations in order to provide more efficient tape operations. We plan to evaluate the proposed strategies on real use cases, and an important effort will be dedicated to reproducibility and open-source development.

The third objective is to improve the integration of the tape storage layer in the HSM.

We plan to combine and improve both operating systems' and tape devices' operations in order to optimize the overall data throughput. This task designs a new piece of software able to proactively optimize and reorder the requests to the tape device: packing data on an application-based, pre-fetching, data-buffering, and write-behind etc. As part of this objective, we will design new data structures to optimize the storage of I/O data on tapes, as well as identifying points of interest to be extracted from the operating system.

Funding & Contacts

This PhD thesis is funded by the ANR JCJC VHS project. The project will benefit of a dedicated hardware setup: a LTO-10 tape drive with associated LTO-10 cartridges, the last standard of the LTO consortium. We will use this drive with LTFS [4], an open-source filesytem for tapes exposing the tape as a disk would be.

The futur PhD student will join the Inria Benagil team, part of the SAMOVAR laboratory of Telecom SudParis & Institut Polytechnique de Paris.

- Supervisor : François TRAHAY (Télécom SudParis)
- Co-advisor : Valentin DELIS (ensIIE)

References

[1] M. C. Davis, V. Bahyl, G. Cancio, E. Cano, J. Leduc, and S. Murray, "CERN Tape Archive – from Development to Production Deployment," in *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics*, ser. EPJ Web of Conferences, vol. 214. EDP Sciences, 2019, p. 04015. [Online]. Available: https://doi.org/10.1051/epjconf/201921404015

[2] M. Mäsker, L. Nagel, T. Süß, A. Brinkmann, and L. Sorth, "Simulation and Performance Analysis of the ECMWF Tape Library System," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Salt Lake city, UT, Nov. 2016, pp. 252–263. [Online]. Available: https://doi.org/10.1109/SC.2016.21

[3] T. Pitakrat, A. Van Hoorn, and L. Grunske, "A comparison of machine learning algorithms for proactive hard disk drive failure detection," in *Proceedings of the 4th international ACM Sigsoft symposium on Architecting critical systems*, 2013, pp. 1–10.

[4] T. Chakravarty, "LTFS Transforms LTO Tape into Nearline Storage: Accelerating 4K Media Workflows," in *SMPTE 2014 Annual Technical Conference & Exhibition*, 2014, pp. 1–7. [Online]. Available: https://doi.org/10.5594/M001577