

1. 担当 PM

田中 邦裕（さくらインターネット株式会社 代表取締役社長）

2. クリエータ氏名

重光 史也

（島根大学 大学院総合理工学研究科 総合理工学専攻 情報システム学コース）

3. 委託金支払額

2,304,000 円

4. テーマ名

NVDIMM 向けファイルシステムの開発

5. 関連 Web サイト

GitHub : <https://github.com/4ge32/aeon>

6. テーマ概要

コンピュータアーキテクチャにおいて、CPU とメモリ、ハードディスクなどの二次記憶の関係性は長年変化がなかったが、不揮発性メモリを活用したストレージクラスのデバイスの登場によって、その関係性に変化が生まれている。そのような中で、メインメモリの空間にフラッシュメモリをマッピングするという NVDIMM の登場は、ある意味コンピュータアーキテクチャを根本的に変えるものであり、その将来性は大きな期待を持たれている。

本プロジェクトは、NVDIMM を利用するための専用ファイルシステムと Linux のカーネルドライバを開発するものである。

7. 採択理由

いま、さまざまな場面において、コンピューティングのアーキテクチャが見直されている。その中でも、不揮発性メモリを活用したアーキテクチャは、今までとは全く異なる価値をもたらすと考えられている。NVDIMM は、既存のコンピュータと相性のいいものではあるが、OS 側からのサポートがまだまだ足りてい

ない状況にあった。今回のプロジェクトは、ファイルシステムを通じて NVDIMM のパフォーマンスを気軽に活用できるという実用性の高いものであり、NVDIMM の普及促進における一助になればと考えていた。また、今までに無いものをカーネルレベルで開発するという心意気が未踏的であることや、波及効果が高いものと考え、採択した。

8. 開発目標

NVDIMM は不揮発性であるという特徴からデータを保存するための記憶媒体としての利用が想定されることが多い。ファイルシステムを開発する意義は、個々のソフトウェア資産の特別なチューニング無しにファイルシステムを入れ替えるだけで NVDIMM の恩恵を享受できるシステムことにある。さらに、NVDIMM を記憶媒体として利用する際ファイルを扱う全てのソフトウェアの基盤となるカーネルのファイルシステムが NVDIMM の特性を活かしたものであるかどうかは、より性能の高いコンピュータシステム実現のために重要な要素となる。

この基本ソフトウェアであるファイルシステムの開発によって、コンピュータ全体の性能の向上を目指した。

9. 進捗概要

本プロジェクトにおいては、プロジェクト名を AEON と称し、既存のファイルシステムに比べて、高速性と透過圧縮機能に特徴を持たせ、実装を進めた。

具体的には、以下の 3 点を中心に実装を行った。

- システムコールの実装
- Direct Access for Files (DAX) を使ったメモリの読み書き
- 透過圧縮機能

システムコールの実装にあたっては、通常の利用者アプリケーションが稼働させることのできるよう、必要とされるシステムコールのすべてに対応するよう、ひとつずつ実装を進めた。

当初は、実装のできていないシステムコールがあり、AEON 上でコンパイルが通らない事象が発生したが、最終的には必要なシステムコールをすべて実装した。DAX については、ファイルの読み書き時に発行される read/write システムコールにおいて、カーネルのページキャッシュをバイパスし NVDIMM へ直接書き込むために使用した。また、mmap システムコール時は、NVDIMM 領域を直接ユーザ空間からアクセスさせるようになっている。AEON では、カーネル API を用いた DAX 読み書きとは別に、NVDIMM 領域をカーネルアドレス空間上において管理している設計を活かした独自 DAX 読み書き関数も備えた。透過圧縮

機能も実装し、アルゴリズムは `zstd` を利用した。バイトアクセス可能な NVDIMM の特徴を活かすような設計を目指し、オーバーヘッドの少ない透過圧縮を実現している。

実装したプロダクトのベンチマークのために、NUMA 構成の NVDIMM-N2 基を搭載した HPE ProLiant DL360 Gen10 を用いた。結果は、マイクロベンチマークで最大 224%、マクロベンチマークで最大 471%、既存のファイルシステムよりも高い性能を示した。比較したファイルシステムは、元々はハードディスク向けであったが NVDIMM 向けのモードでの対応がなされた EXT2, EXT4, XFS, さらに NVDIMM 向けにフルスクラッチされたファイルシステムである NOVA と比較している。いずれのファイルシステムに対しても AEON は速度性能で優れた数値を示した。

10. プロジェクト評価

本プロジェクトにおいては、特に有用ではないと考えた `mmap` の実装を除いて、当初の想定通りに実装を完了させることができ、ファイルシステムとしての動作も確認できた。プロジェクト初期においては、想定機能が利用できないなどの不具合もあったが、実装が進むにつれて、今まで動かなかったアプリケーションが当ファイルシステム上で稼働するようになり、プロジェクトのプロセスと成果を感じることでできるプロジェクトであった。実装上においては、NUMA (Non-Uniform Memory Access) への対応や、フリーリストの効率的な管理など、単に動作するだけではなく、堅牢性や高速性なども加味しながら、実用に耐えうるレベルでのコーディングが出来た。加えて、ファイルシステムの透過的な圧縮など、競合するファイルシステムにはない付加価値的な機能の実装も行えたことから、今後も開発が継続するのであれば、一般的に利用されるものになると考える。

ちなみに、本プロジェクトの開発においては NVDIMM を搭載したサーバの実機がなく、KVM 上でのシミュレーション環境での開発であったが、日本ヒューレット・パカード社の好意により実機を借り受けることができ、実機上での性能検証も実施した。

結果としては、競合するファイルシステムよりも高い性能を確認でき、本プロジェクトの成果物の有効性も確認することができたことから、当初の想定を超えるアウトプットになったと考える。

11. 今後の課題

本プロジェクトの成果はオープンソースソフトウェアとして GitHub 上に公開されており、現在もバグ修正や機能追加、改善を行っている。ファイルシステムにおいては、完全性を強く求められるものであり、さらなる信頼性を高めていくことが必要である。また、オープンソースソフトウェアとして認知を引き上げ

ていくことが今後の課題である.