

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Nghiên cứu các kỹ thuật đệm (caching) cho SSD hiệu năng cao

NGUYỄN TRUNG ĐỨC

duc.nt204725@sis.hust.edu.vn

Ngành: Kỹ thuật máy tính

Chuyên ngành: Kỹ thuật máy tính

Giảng viên hướng dẫn: TS. Đỗ Công Thuần

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2024

LỜI CẢM ƠN

Trong suốt quá trình thực hiện đồ án tốt nghiệp với đề tài “Nghiên cứu các kỹ thuật đệm (caching) cho SSD hiệu năng cao”, em đã nhận được sự giúp đỡ, đóng góp ý kiến và chỉ bảo nhiệt tình từ các thầy cô, gia đình và bạn bè. Em xin bày tỏ lòng biết ơn sâu sắc đối với mọi sự hỗ trợ này.

Lời đầu tiên, em xin được gửi lời cảm ơn chân thành tới giáo viên hướng dẫn của mình, TS. Đỗ Công Thuần, vì sự hướng dẫn quý báu, sự hỗ trợ và khích lệ trong suốt quá trình em hoàn thành đồ án tốt nghiệp. Những góp ý, nhận xét và chỉ bảo của thầy đã giúp em vượt qua những khó khăn trong quá trình thực hiện.

Em cũng xin bày tỏ lòng biết ơn đến các thầy cô tại Trường Công nghệ Thông tin và Truyền thông. Sự chỉ dạy và hướng dẫn của thầy cô trong suốt quá trình học tập đã giúp em xây dựng được nền tảng kiến thức vững chắc, là tiền đề quan trọng để hoàn thành quá trình nghiên cứu đồ án này.

Cuối cùng, em muốn bày tỏ lòng biết ơn sâu sắc nhất của mình đến gia đình và bạn bè vì đã giúp đỡ, ủng hộ em không chỉ trong quá trình hoàn thành đồ án tốt nghiệp này mà còn suốt thời gian học tại Đại học Bách Khoa Hà Nội.

Cảm ơn Bách Khoa!

LỜI CAM KẾT

Họ và tên sinh viên: Nguyễn Trung Đức

Điện thoại liên lạc: 0779210768

Email: duc.nt204725@sis.hust.edu.vn

Lớp: Kỹ thuật máy tính 01 - K65

Hệ đào tạo: Kỹ thuật máy tính

Tôi – *Nguyễn Trung Đức* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS. Đỗ Công Thuần*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày 01 tháng 01 năm 2025

Tác giả ĐATN

Nguyễn Trung Đức

TÓM TẮT NỘI DUNG ĐỒ ÁN

Trong kỷ nguyên số hóa, nhu cầu xử lý và lưu trữ dữ liệu tăng trưởng mạnh mẽ, đặc biệt trong các lĩnh vực như trí tuệ nhân tạo, phân tích dữ liệu lớn, và điện toán đám mây. SSD (Solid-State Drive) đang trở thành lựa chọn lưu trữ ưu việt nhờ tốc độ truy xuất cao, tiêu thụ năng lượng thấp và độ tin cậy vượt trội. Tuy nhiên, SSD cũng đối mặt với các hạn chế như số lần ghi/đọc hữu hạn và chi phí cao, đòi hỏi các giải pháp tối ưu để cải thiện hiệu năng và kéo dài tuổi thọ thiết bị. Điều này khiến nghiên cứu về các kỹ thuật đệm (caching) trở nên quan trọng, nhằm cân bằng giữa tốc độ xử lý và độ bền SSD, đáp ứng tốt hơn các yêu cầu hiện đại.

Nghiên cứu về các kỹ thuật đệm cho SSD hiệu năng cao tập trung vào việc tối ưu hóa truy xuất dữ liệu và giảm tải cho thiết bị bằng các phương pháp như đệm ghi, đệm đọc và đệm lai. Các thuật toán quản lý bộ nhớ đệm như LRU, LFU và ARC được áp dụng để quản lý hiệu quả dữ liệu trong bộ nhớ đệm. Bên cạnh đó, các giải pháp giảm khuếch đại ghi, cân bằng hao mòn và tích hợp bộ nhớ NVRAM giúp cải thiện hiệu suất tổng thể và kéo dài tuổi thọ SSD. Những kỹ thuật này không chỉ giúp tối ưu hóa thiết bị cá nhân mà còn nâng cao hiệu quả hoạt động của các hệ thống lưu trữ lớn, đáp ứng nhu cầu ngày càng cao về tốc độ và độ tin cậy trong môi trường công nghệ hiện đại.

ĐATN tìm hiểu về SSD hiện đại, nghiên cứu các kỹ thuật đệm cho SSD hiệu năng cao và đề xuất thuật toán giảm truy cập vào SSD, giúp tăng tuổi thọ cho SSD hiện đại. ĐATN đã triển khai mô phỏng khả thi bằng trình mô phỏng MQSim và có kết quả khá tốt.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Hệ thống lưu trữ SSD hiện đại và vấn đề của SSD hiện nay	2
1.1.1 Tổng quan về hệ thống lưu trữ	2
1.1.2 Sự phát triển của hệ thống lưu trữ SSD.....	2
1.1.3 Những công nghệ lưu trữ SSD và vấn đề hiện nay của SSD.....	3
1.2 Giới thiệu tổng quát về ĐATN	4
1.2.1 Mục tiêu của đề án.....	4
1.2.2 Phương pháp nghiên cứu.....	5
1.2.3 Đóng góp của đề án.....	5
1.3 Bố cục đề án	5
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT	7
2.1 Các lý thuyết nền tảng trong hệ thống lưu trữ SSD.....	7
2.1.1 Nguyên lý bộ nhớ Flash NAND.....	7
2.1.2 Lý thuyết bộ nhớ đệm và quản lý dữ liệu	9
2.1.3 Cấu trúc và giao thức kết nối.....	9
2.1.4 Lý thuyết bảo vệ và khôi phục dữ liệu	9
2.1.5 Kiến trúc đa kênh (Multi-Channel Architecture)	9
2.2 Vấn đề tuổi thọ của SSD hiện nay.....	10
2.3 Một số thuật toán trong hệ thống SSD	10
2.3.1 Thuật toán Least Recently Used (LRU)	10
2.3.2 Thuật toán Least Frequently Used (LFU)	11
2.3.3 Thuật toán Adaptive Replacement Cache (ARC).....	13
2.3.4 Thuật toán Write Amplification Reduction	14
2.3.5 Thuật toán Hot and Cold Data Separation.....	15

CHƯƠNG 3. MÔ HÌNH HỆ THỐNG.....	18
3.1 Cấu trúc bên trong SSD.....	18
3.1.1 Logic giao diện máy chủ (HIL).....	19
3.1.2 Lớp chuyển đổi flash (FTL)	19
3.2 Mô hình được sử dụng trong ĐATN	20
3.2.1 Mô hình Back-End SSD.....	21
3.2.2 Mô hình Front-End SSD	22
3.3 Mô hình triển khai	23
3.3.1 Thuật toán đệm	24
3.3.2 Mô-đun giám sát tải cấp die	24
3.3.3 Mô-đun ghi	24
3.4 Cấu trúc source code của ĐATN	24
3.4.1 Thư mục <i>src</i>	24
3.4.2 Thư mục <i>traces</i>	37
3.4.3 File <i>config</i>	38
CHƯƠNG 4. THUẬT TOÁN ĐỀ XUẤT.....	39
4.1 Cách thức hoạt động	39
4.2 Quản lý dữ liệu	41
4.3 Mô-đun giám sát tải cấp die.....	41
4.4 Triển khai thuật toán	43
CHƯƠNG 5. MÔ PHỎNG VÀ ĐÁNH GIÁ KẾT QUẢ	49
5.1 Cài đặt hệ thống	49
5.1.1 Yêu cầu hệ thống.....	49
5.1.2 Cài đặt và triển khai	49
5.2 Tùy chỉnh cấu hình.....	49

5.3 Các tham số mô phỏng	50
5.3.1 Host Parameter Set.....	50
5.3.2 Device Parameter Set	50
5.3.3 Flash Parameter Set.....	52
5.4 Kết quả chạy thuật toán và so sánh	53
5.4.1 Input/Output Operations Per Second	53
5.4.2 Bandwidth.....	54
5.4.3 Flash Command.....	54
5.4.4 Tỷ lệ Hit/Miss Cache	55
5.5 Phân tích kết quả	56
CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	57
6.1 Kết luận.....	57
6.2 Hướng phát triển trong tương lai	57
TÀI LIỆU THAM KHẢO.....	59

DANH MỤC HÌNH VẼ

Hình 1.1	Mô hình máy tính hiện đại	1
Hình 2.1	Tổ chức của bộ nhớ Flash NAND	7
Hình 2.2	Minh họa thuật toán LRU	11
Hình 2.3	Minh họa thuật toán LFU	12
Hình 2.4	Minh họa thuật toán ARC	13
Hình 2.5	Minh họa thuật toán Write Amplification Reduction	14
Hình 2.6	Minh họa thuật toán Hot and Cold Data Separation	16
Hình 3.1	Cấu trúc của một MQ-SSD.	19
Hình 3.2	Tổng quan về các thành phần của MQSim	21
Hình 3.3	Mô hình tổng quan	23
Hình 5.1	Kết quả mô phỏng chỉ số IOPS cho 3 workloads của thuật toán đề xuất(DATN) so với chuẩn (Base)	53
Hình 5.2	Kết quả mô phỏng chỉ số Bandwidth cho 3 workloads của thuật toán đề xuất(DATN) so với chuẩn (Base)	54
Hình 5.3	Kết quả mô phỏng số lệnh đọc flash cho 3 workloads của thuật toán đề xuất(DATN) so với chuẩn (Base)	54
Hình 5.4	Kết quả mô phỏng số lệnh ghi flash cho 3 workloads của thuật toán đề xuất(DATN) so với chuẩn (Base)	55
Hình 5.5	Kết quả mô phỏng tỉ lệ Hit/Miss cho 3 workloads của thuật toán đề xuất(DATN) so với chuẩn (Base)	55

DANH MỤC BẢNG BIỂU

Bảng 5.1	Tham số của Host Parameter Set	50
Bảng 5.2	Tham số của Device Parameter Set	51
Bảng 5.3	Tham số của Flash Parameter Set	52