

2026 Winter RocksDB Study Kick-off Meeting

Yongmin Lee, Dayeon Wee

<http://sslabs.dankook.ac.kr/>, <https://sslabs.dankook.ac.kr/~choijm>

Presentation by Yongmin Lee
nascarf16@dankook.ac.kr

Contents

- 1. About the Professor**
- 2. About the Assistants**
- 3. Self-Introduction**
- 4. Why should We Study NoSQL**
- 5. RocksDB Overview**
- 6. Study Schedule**

2026 Winter RocksDB Study

Tasks for Today's Meeting

- ◎ Advisor Introduction
- ◎ Self-Introduction
- ◎ Simple RocksDB Overview
- ◎ Potential Research Topic

Tasks for This Study Session

- ◎ Share Experiences with Open-Source Analysis
- ◎ Manage GitHub Repository (To Be Assigned)
- ◎ Decide on a Study Group Name
- ◎ Write a Research Paper (Final Goal)

오픈 소스 SW 심화 스터디

비정형 빅데이터를 위한 RocksDB 구조 분석과 기술 동향 이해



목표

- 구글, 페이스북, 아마존 등에서 사용하는 빅데이터 관리 SW 이해
- 대표적인 카-벨류 DB인 RocksDB 소스 코드 수준 분석
- LSM-tree, 집약, 검색, 불륨 필터 등 핵심 자료구조 이해
- 실질적이고 전문가 수준의 오픈 소스 분석 및 기여

대상

- 시스템 프로그래밍, 자료구조 수업 이수자 중 위의 내용에 관심 있는 사람.
- 쓰레드, 파일, 동기화, 페이지 캐시 학교에서 배운 내용이 등이 실제 산업계 SW에서 어떻게 사용되는지 관심이 있는 사람.
- 다양한 워크로드에서 실행 및 성능 분석 관심이 있는 사람.
- C 또는 C++를 다룰 수 있으며 위의 내용에 관심있는 전공자/비전공자 및 재학생/졸업생 누구나 참여가능
- 지원 연락처 - 이웅민 (nascart16@ Dankook.ac.kr) 위다연 (wida10@Dankook.ac.kr)

스터디 상세 정보

- | | |
|--------|---|
| 일정 | 2026년 1/5일 ~ 2/27일
(매주 1번씩 세미나 진행, 화요일 오후 2-4시 예정) |
| 장소 | 단국대학교 SW-ICT관 301호 |
| 강사 | 최종무/유시환 교수, 관련 대학원 학생들 |
| 세미나 내용 | |
| 1주차 | 빅데이터 특성, 관리 SW 소개, RocksDB 소스 코드 수준 소개 |
| 2주차 | RocksDB 핵심 구조, LSM-tree, 집약, 검색 |
| 3주차 | RocksDB 최근 연구 동향, Wickey, SILK, Remix 등 |
| 4주차 | RocksDB interface 분석(학생 2-3명 발표, 교수/대학원생 지도) |
| 5주차 | RocksDB compaction 분석(학생 2-3명 발표, 교수/대학원생 지도) |
| 6주차 | RocksDB lookup 분석(학생 2-3명 발표, 교수/대학원생 지도) |
| 7주차 | RocksDB new approach 분석(학생 2-3명 발표, 교수/대학원생 지도) |
| 8주차 | 새로운 기법 구현 및 성능 평가, (가능하면) 논문 작성 |
- 위 내용을 진행함에 따라 변경 가능

1001

4

About the Professor

Prof. Seehwan Yoo



Seehwan Yoo leads the Mobile-OS Lab at Dankook University, guiding students through fundamental and applied problems across operating systems, cloud infrastructure, trusted execution, and storage systems. His group bridges low-level systems design with empirical analysis—covering SGX and TrustZone enclaves, Xen-ARM real-time virtualization, RocksDB/LevelDB on ZNS SSDs, and SIMD-friendly learned index structures.

He emphasizes reproducibility and infrastructure you can actually ship: Kubernetes-based teaching clusters, Ceph + ZNS benches, and SGX exploit harnesses. Yoo's work appears in IEEE TMC, IEEE Access, ACM SoCC, BigComp, NVMSA, SYSTOR, and other international and Korean venues.

Current Focus Areas

- Secure execution environments (SGX, TrustZone, Python JIT hardening)
- Storage systems and tiered compaction on ZNS SSDs, Ceph, RocksDB/BlobDB
- Learned indexes (RMI, ALEX, SIMD-accelerated search, RB-SkipList variants)
- Kubernetes-based education clouds, CryptoJacking observability pipelines

TITLE	CITED BY	YEAR
DAoT: Dynamic and energy-aware authentication for smart home appliances in Internet of Things YP Kim, S Yoo, C Yoo 2015 IEEE International conference on consumer electronics (ICCE), 196-197	58	2015
PARFAIT: A new scheduler framework supporting heterogeneous Xen-ARM schedulers JW Jeong, S Yoo, C Yoo 2011 IEEE Consumer Communications and Networking Conference (CCNC), 1192-1196	58	2011
Eliminating the Performance Anomaly of 802.11 b S Yoo, JH Choi, JH Hwang, C Yoo International Conference on Networking, 1055-1062	51	2005
Method for adjusting a transmission rate to obtain the optimum transmission rate in a mobile ad hoc network environment J Sin, B Mun, S Cho, H Yoo, S Yoo, JH Choi US Patent 7,336,608	42	2008
Spatio-temporal split learning for privacy-preserving medical platforms: Case studies with covid-19 ct, x-ray, and cholesterol data YJ Ha, M Yoo, G Lee, S Jung, SW Choi, J Kim, S Yoo IEEE Access 9, 121046-121059	34	2021
Mobivmm: a virtual machine monitor for mobile phones S Yoo, Y Liu, CH Hong, C Yoo, Y Zhang Proceedings of the First Workshop on Virtualization in Mobile Computing, 1-5	32	2008
Feasibility study of multi-site split learning for privacy-preserving medical systems under data imbalance constraints in COVID-19, X-ray, and cholesterol dataset YJ Ha, G Lee, M Yoo, S Jung, S Yoo, J Kim Scientific Reports 12 (1), 1534	30	2022
Confzns: A novel emulator for exploring design space of zns ssds I Song, M Oh, BSJ Kim, S Yoo, J Lee, J Choi Proceedings of the 16th ACM International Conference on Systems and Storage ...	25	2023
A case for bad big. LITTLE switching: how to scale power-performance in SI-HMP S Yoo, YS Shim, S Lee, SA Lee, J Kim Proceedings of the Workshop on Power-Aware Computing and Systems, 1-5	21	2015

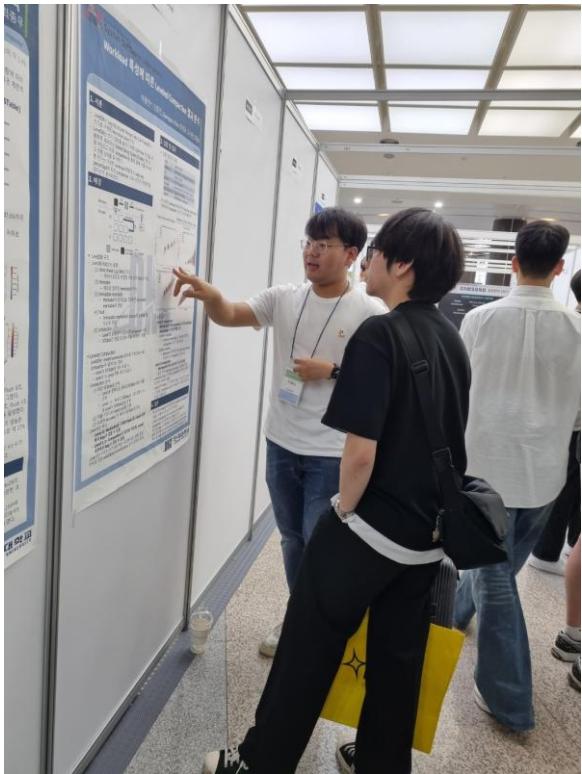
<https://scholar.google.com/citations?user=jLjV6bwAAAAJ&hl=en&oi=ao>

About the Assistants

Assistant - 이용민



<input type="checkbox"/> 내서재 담기	Excel 다운	최신순 ▼
<input type="checkbox"/> Workload 특성에 따른 Leveled Compaction 결과 분석		
이용민, 신호진, Guangxun Zhao Korean Institute of Information Scientists and Engineers 한국정보과학회 학술발표논문집 2025.07 2509-2511 (3pages) 학술대회자료 이용수 5		
<input type="checkbox"/> LSM-tree 최적화를 위한 레벨 기반 SSTable 크기 조정 기법		
Guangxun Zhao, Yongjie Zhu, 이용민 Korean Institute of Information Scientists and Engineers 한국정보과학회 학술발표논문집 2025.07 1703-1705 (3pages) 학술대회자료 이용수 13		
<input type="checkbox"/> From Distance to Distribution: 고차원 벡터 탐색을 위한 확률 기반 클러스터 선택 기법 연구		
최연규, Guangxun Zhao, Charles Jaranilla Korean Institute of Information Scientists and Engineers 한국정보과학회 학술발표논문집 2025.07 251-253 (3pages) 학술대회자료 이용수 8		



About the Assistants

Assistant - 위다연



위다연

단국대학교
dankook.ac.kr의 이메일 확인됨

- ☐ 제목
- ☐ **키-밸류 크기 비율이 RocksDB 성능에 미치는 영향 분석**
위다연, 신호진, 유시환, 최종무
한국정보과학회 학술발표논문집, 1697-1699
- ☐ **캐시 정책 조합이 성능 지표에 미치는 영향 분석**
최건희, 신호진, 이성현, 위다연, 이재동, 유시환, 최종무
한국정보과학회 학술발표논문집, 1769-1771

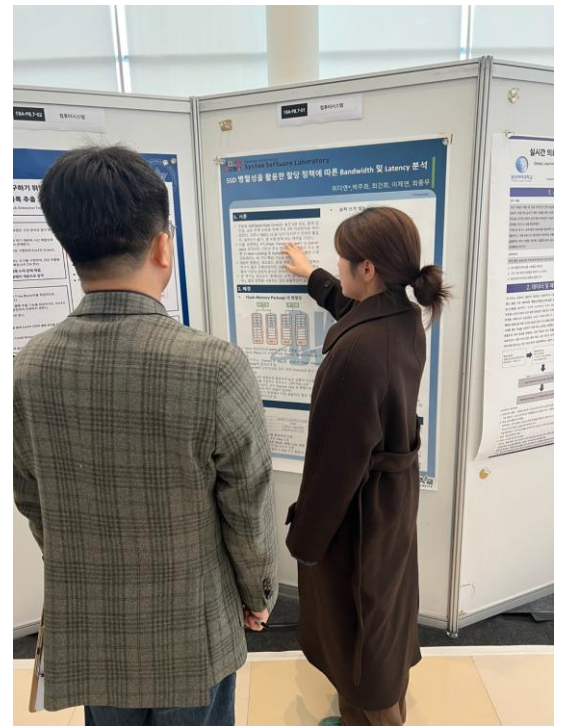


[06704] 서울시 서초구 방배로 76 (방배동, 머리재빌딩 401호) T.1588-2728 F.02-521-1352 E-mail:kiise@kiise.or.kr

제25-1130호

정보과학회 논문지 우수논문 초청

단국대학교 위다연 님께



Self-Introduction



Why Should We Study NoSQL

NoSQL (Not only SQL)

- ◎ Structured data: well defined format, schema-oriented
 - ◎ Table, Row (Record, Tuple), Column (Field, Attribute), Value
- ◎ **Unstructured data**: various and diverse, schema-free
 - ◎ Various fields/frequencies in each record, Diverse format, ...

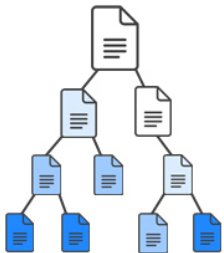
id	name	age
1	Jim	28
2	Pam	26
3	Michael	42

id	subject	Teacher
1	Languages	John Jones
2	Track	Wally West
3	Swimming	Arthur Curry
4	Computers	Victor Stone

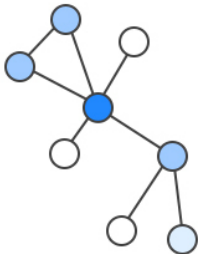
student_id	subject_id	grade
2	1	98
1	2	100
1	4	75
3	3	60
2	4	76
3	2	88

→ **NoSQL** DB: for unstructured data (e.g., LevelDB, RocksDB)

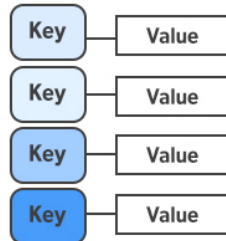
Document



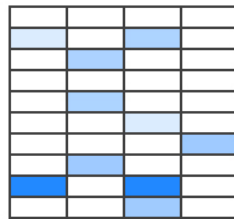
Graph



Key-Value



Wide-column



Google

- Bigtable, Level DB, Hbase
- For Web indexing and messaging

Amazon

- Dynamo, SimpleDB
- For E-commerce

ORACLE

Oracle

- NoSQL, Berkeley DB
- For Configurable I

amazon.com

Microsoft

- Azure, Cosmos DB
- For E-commerce

Facebook

- Haystack, RocksDB, Cassandra
- For social network and photo store

Baidu

- Atlas
- For Cloud data

Yahoo

- PNUTS
- For Advertising

LinkedIn

- Voldemort
- For Scalability

Basho

- Riak
- For distributed KV

Open source

- Redis, Memcached
- For in-memory DB, cache

Why Should We Study NoSQL

Preferred Qualifications for the Company

LINE CAREERS People Culture Jobs FAQs

담당업무

- MongoDB 클러스터의 운영 관리
- MongoDB 클러스터의 관리 도구 개발 및 유지/보수
- MongoDB 관련 신기술 도입 검토 및 연구
- 클라우드 플랫폼에 MongoDB 제품 기획

자격요건

- 1가지 이상의 RDBMS 또는 NoSQL 제품을 2년 이상 운영 관리한 경험이 있으신 분
- 1년 이상 LINUX 환경에서 개발/운영 업무를 진행해 보신 분
- 1년 이상 개발 경험이 있으신 분 (Python 또는 Go)
- 매사 긍정적이고 책임감이 있으며 능동적인 분
- 협업 및 커뮤니케이션이 원활하신 분
- 해외 출장 및 근무에 결격 사유 없으신 분

우대사항

- 2년 이상의 MongoDB 구축 및 운영 경험이 있으신 분
- Rundeck 또는 Ansible을 사용하여 자동화에 대한 구축 경험이 있으신 분
- CI/CD 구축 경험이 있으신 분
- Open Source 도입 및 경험과 관심이 많으신 분

Line

Kakao

주요업무

- 하이브리드 멀티 클라우드(On Premise 포함) 환경에서 Database 관리와 톨 구축
- No-SQL Database 에 대한 설계와 관리
- No-SQL Database 운영 및 자동화
- No-SQL Database 백업 및 복구
- No-SQL Database Data Handling
- No-SQL Database High Availability
- MS-SQL Database 상품 기획 및 운영 관리
- 오픈소스 No-SQL 데이터베이스의 리뷰 및 적용 방안 검토
- 클라우드 환경에 적합한 No-SQL Database 기반 서비스 설계와 개발

자격요건

[지원자격]

- No-SQL(Redis, MongoDB, HBase etc) Database의 설계와 운영관리 경험
- 프로그래밍 개발 및 스크립트 작성이 가능하신 분
- 새로운 기술에 대한 거부감이 없고 도전적인 분
- 논리적이고 원활한 커뮤니케이션 역량을 보유하신 분

11번가(실일번가) · 서울 · 경력 3-20년

NoSQL DBA

포지션 상세

고객으로부터 신뢰 받는 최고의 Commerce Portal, 11번가! 11번가에서 함께 성장할 동료를 기다리고 있습니다.

주요업무

- 11번가 서비스에 관련된 MongoDB 운영 및 관리
- DBMS 운영 효율화 및 자동화
- NoSQL DB Performance Tuning

자격요건

- NoSQL 구축 및 운영 경험 3년 이상하신 분
- NoSQL 이중화 구성 및 백업/복구, 트러블슈팅 경험이 있으신 분
- NoSQL 스키마 관리 및 튜닝 경험

RocksDB overview

(1) What is RocksDB?



- ◎ **Key-Value Store:** Stores Key-Value data (like Python Dict)
- ◎ **Write Optimized Database: LSM-Tree** Structure
- ◎ Using **Out-of-Place Update**
- ◎ Forked from Google's **LevelDB**, developed by Meta (FaceBook)
- ◎ Support various algorithms, configurations, tools and debugging facilities

Dictionary in Python PYnative.com

Unordered collections of unique values stored in (Key-Value) pairs.

```
d = {'a': 10, 'b': 20, 'c': 30}
```

↑
d['a']

↑
d['b']

↑
d['c']

- ✓ **Unordered:** The items in dict are stored without any index value
- ✓ **Unique:** Keys in dictionaries should be Unique
- ✓ **Mutable:** We can add/Modify/Remove key-value after the creation

google/leveldb

LevelDB is a fast key-value storage library written at Google that provides an ordered mapping from string keys to string...



57
Contributors

22
Used by

39k
Stars

8k
Forks



facebook/rocksdb

A library that provides an embeddable, persistent key-value store for fast storage.



791
Contributors

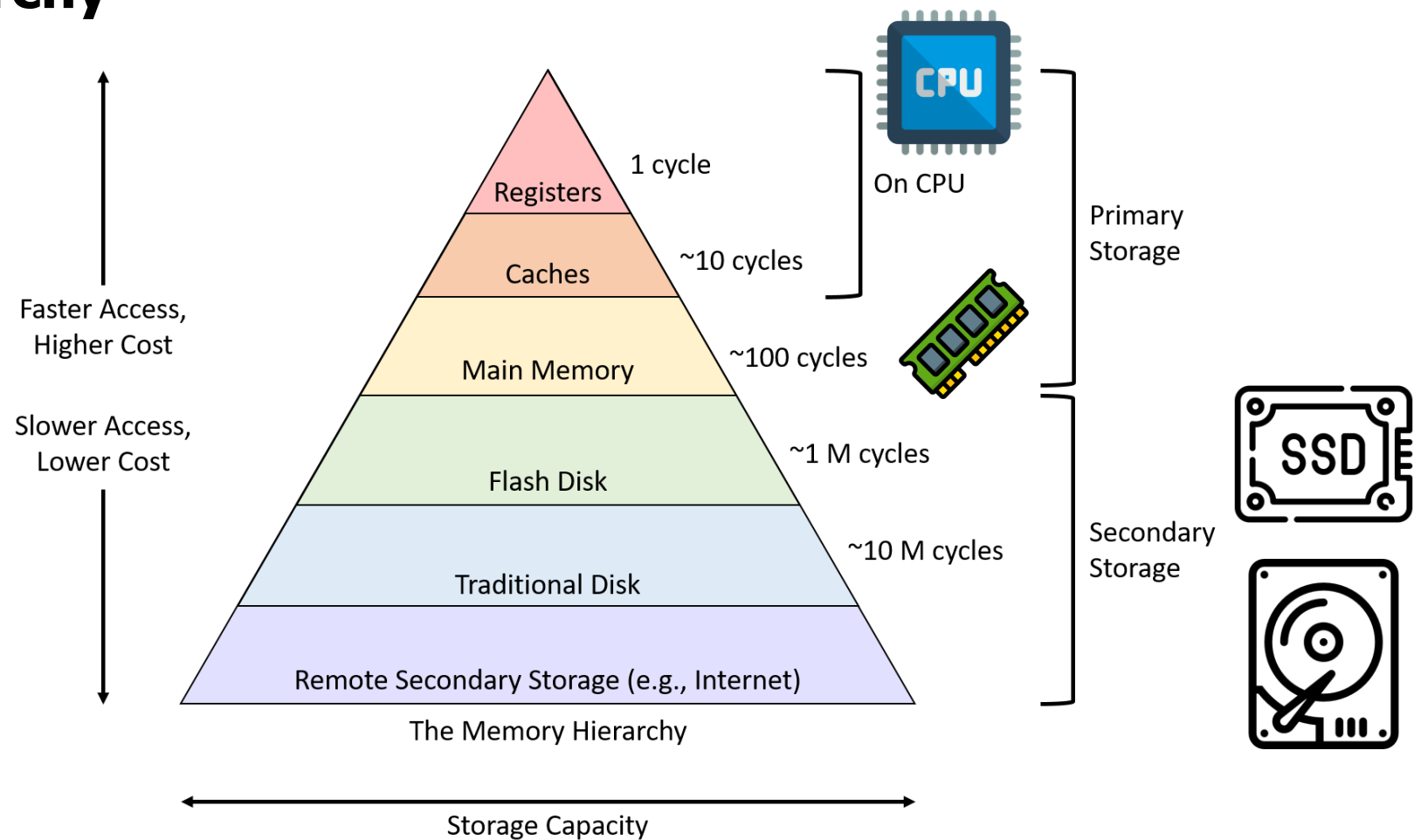
20
Used by

31k
Stars

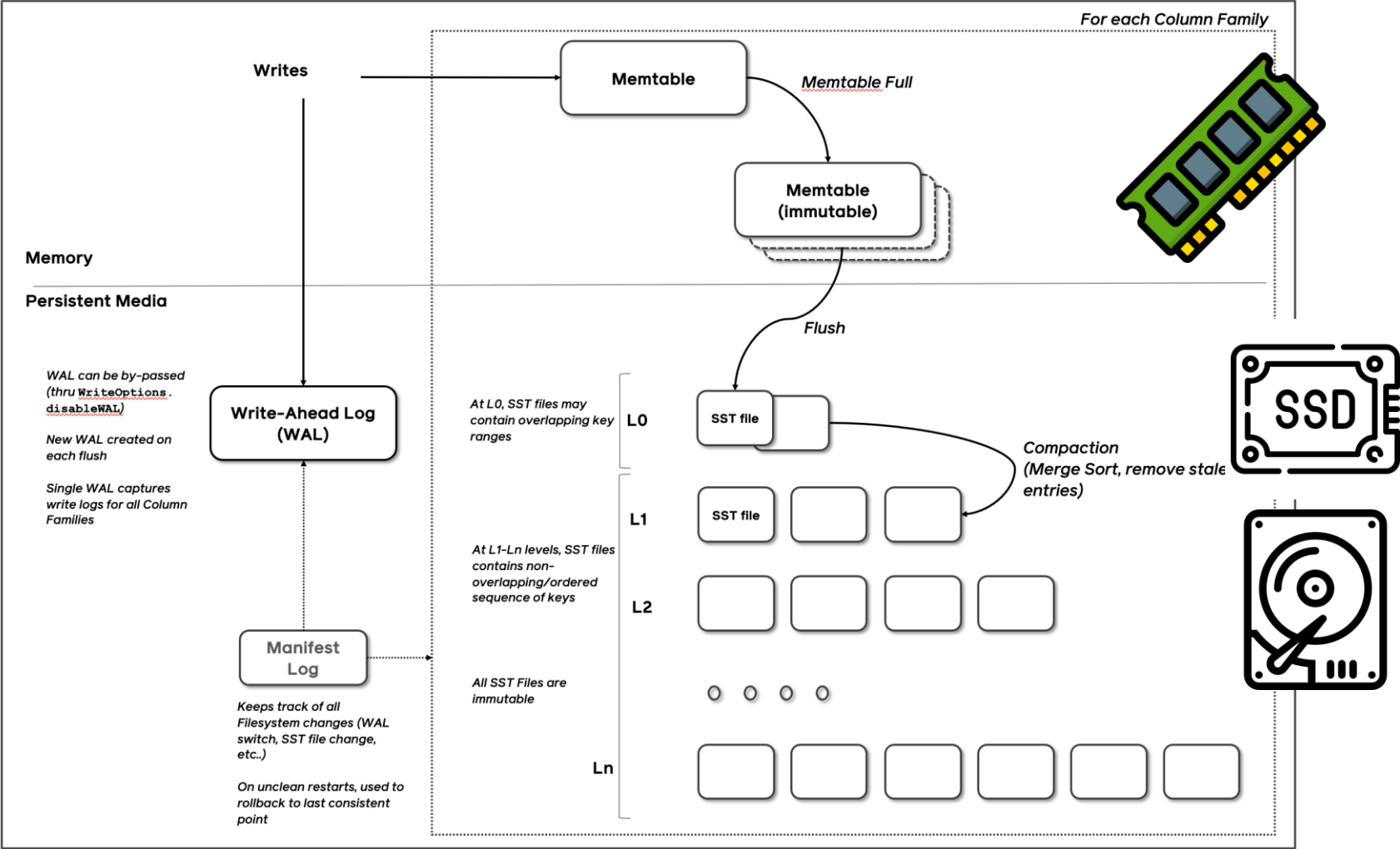
7k
Forks



(2) Memory Hierarchy



(3) RocksDB Overview

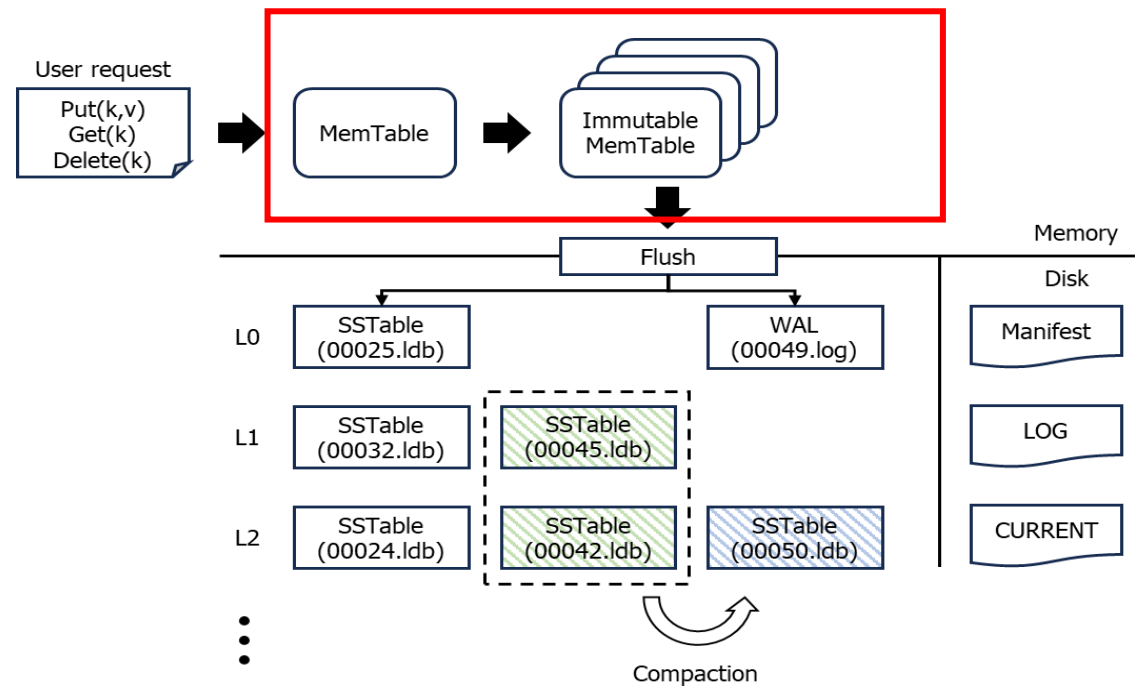


- ◎ **FAST**
- ◎ **EXPENSIVE**
- ◎ **CHEAP**
- ◎ **SLOW**

Potential Research Topic

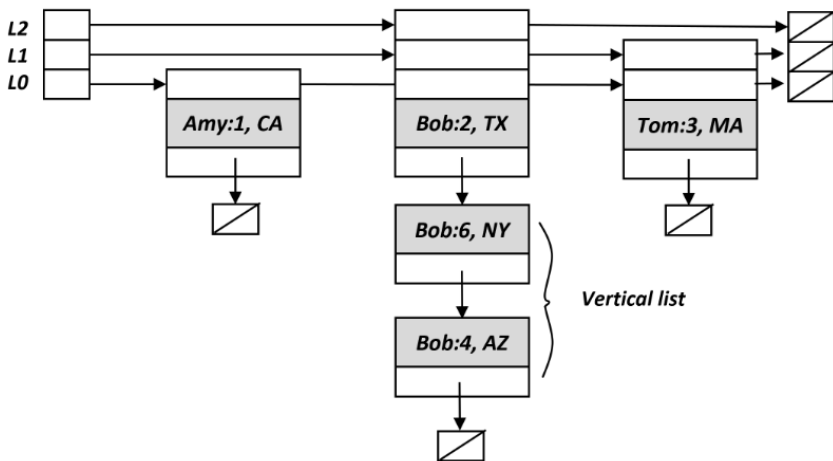
(1) In-memory Index Optimization

- © SkipList (LevelDB, RocksDB, Redis)
- © Hash Table (Memcached, Redis)



LevelDB, RocksDB Architecture

JellyFish (Middleware'20)



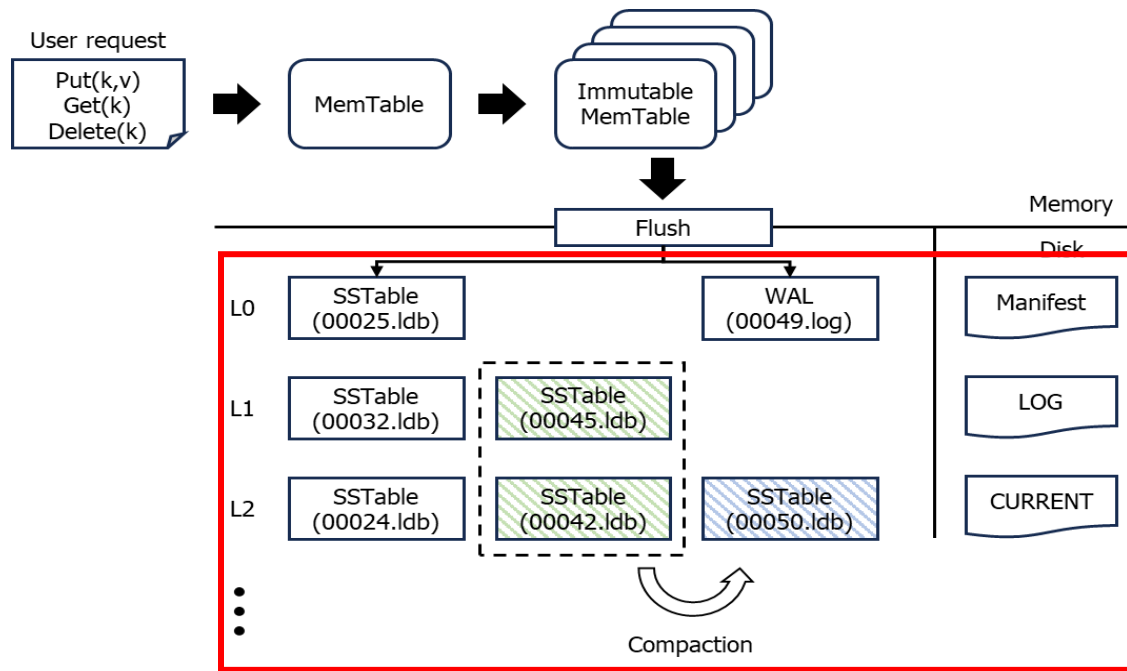
8 Data Structures That Power Your Databases ByteByteGo.com

Types	Illustration	Use Case	Note
Skiplist		In-memory	used in Redis
Hash index		In-memory	Most common in-memory index solution
SSTable		Disk-based	Immutable data structure. Seldom used alone
LSM tree		Memory + Disk	High write throughput. Disk compaction may impact performance
B-tree		Disk-based	Most popular database index implementation
Inverted index		Search document	Used in document search engine such as Lucene
Suffix tree		Search string	Used in string search, such as string suffix match
R-tree		Search multi-dimension shape	Such as the nearest neighbor

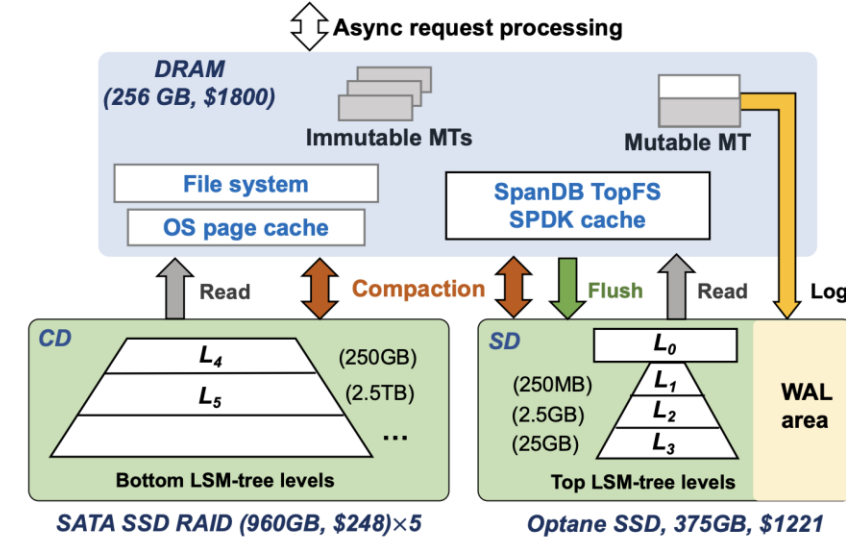
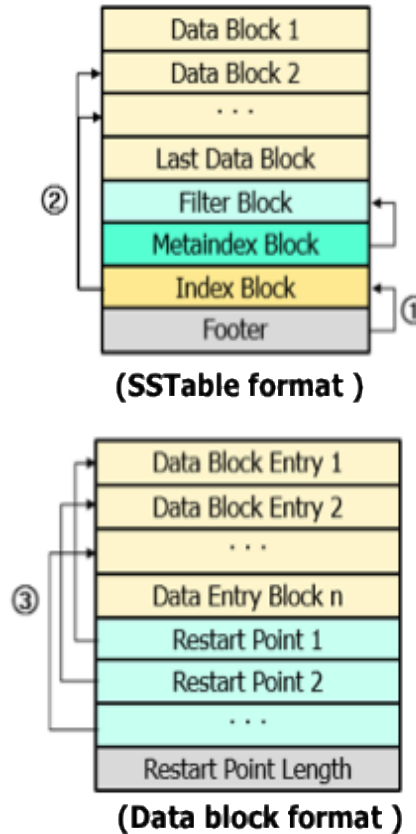
Potential Research Topic

(2) Storage Optimization

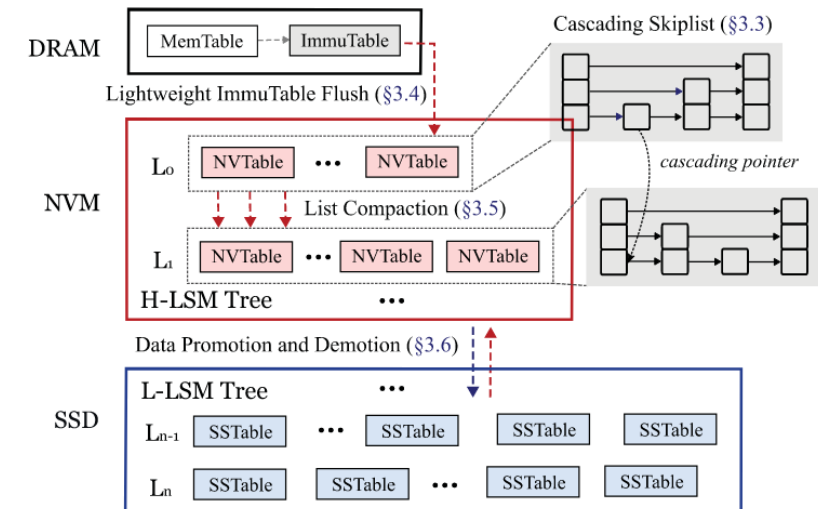
- ◎ Tiered Storage (Slow / Fast Disks)
- ◎ Non-Volatile Memory



LevelDB, RocksDB Architecture



SpanDB (FAST'21)

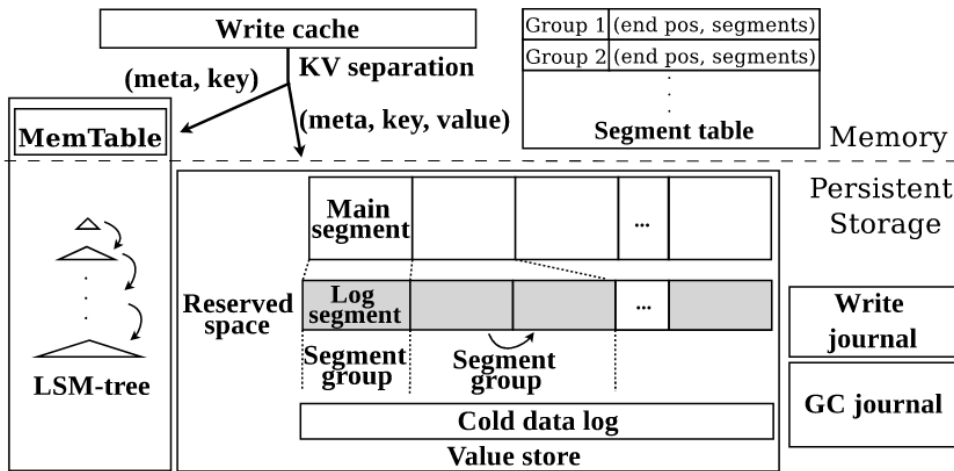


SplitDB (IEEE ToC'24)

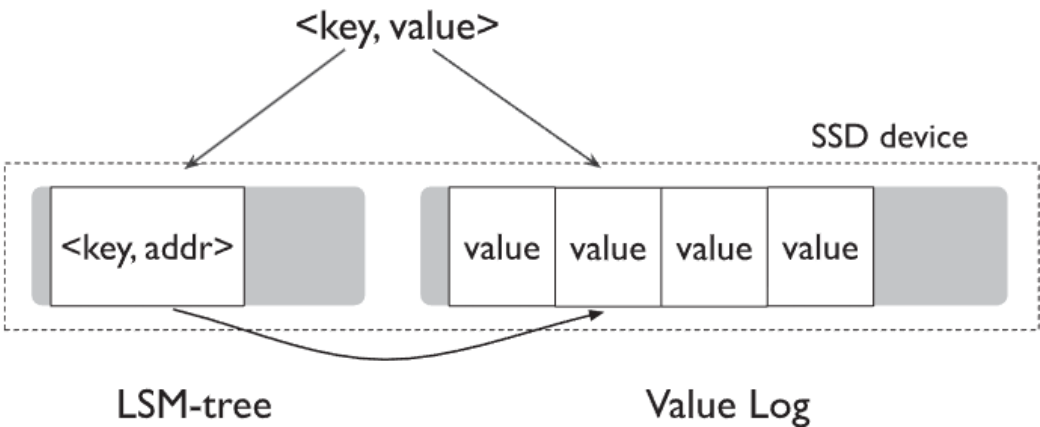
Potential Research Topic

(3) KV Separation Mechanism

- ◎ For handle Large Key-Value size
 - ◎ WiscKey (FAST'16): Pioneer
 - ◎ HashKV (ATC'18): Hash Based Data Grouping
 - ◎ DiffKV (ATC'21): Various Key-Value Size



HashKV (ATC'18)



WiscKey (FAST'16)

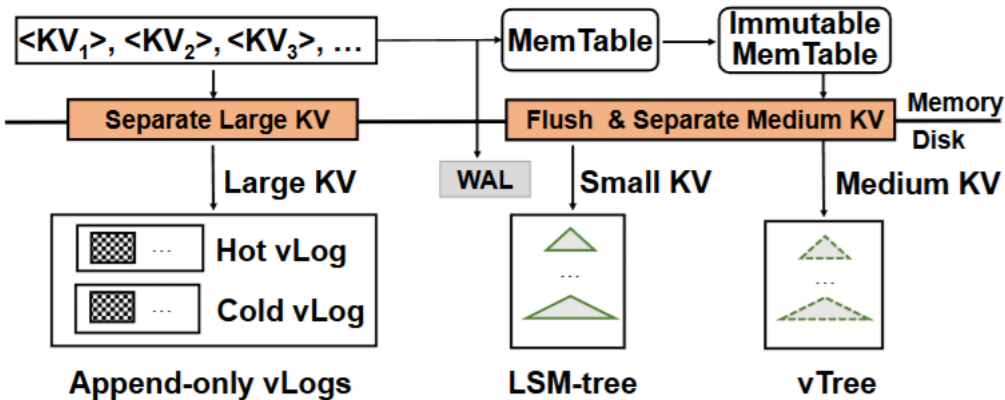


Figure 8: Fine-grained KV separation.
DiffKV (ATC'21)

Potential Research Topic

(4) Learned Index

- ◎ ML for System
- ◎ Learning Unit
 - ◎ SSTable (Bourbon, LeaderKV)
 - ◎ Value Log (LearnedKV)

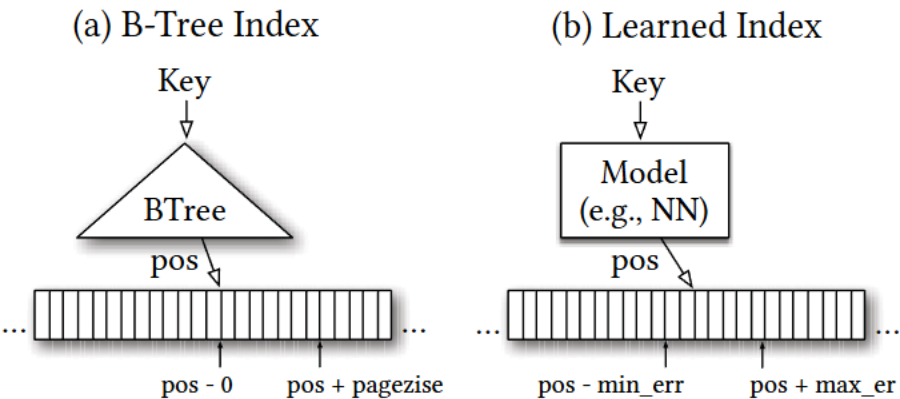
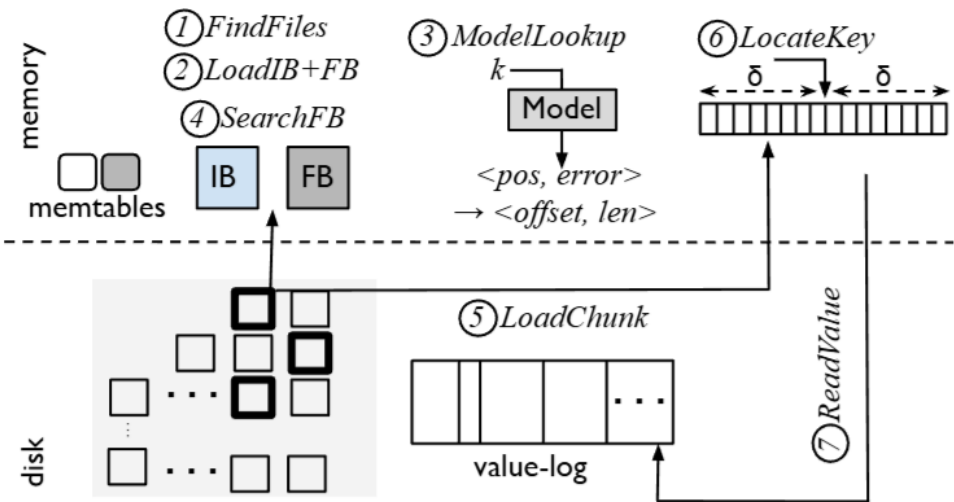


Figure 1: Why B-Trees are models

Learned Index (SIGMOD'18)



Bourbon (OSDI'20)

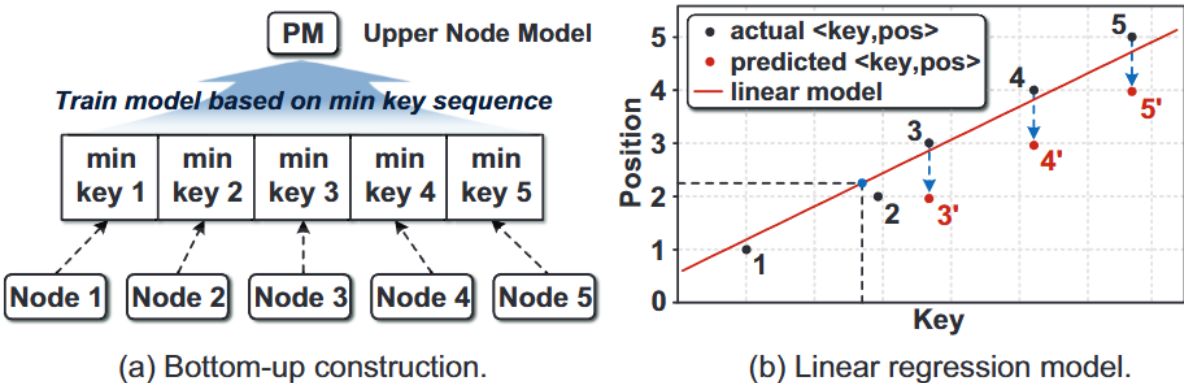


Fig. 6: An example to illustrate the redirect mechanism.

LeaderKV (ICDE'24)

Study Schedule

Preferred day of the week

- 1. Monday
- 2. Tuesday
- 3. Wednesday

Theory-based Lecture

Presentation-based Lecture

Final Presentation !!!

일	월	화	수	목	금	토
28일	29일	30일	31일	1월 1일	2일	3일
Jan			★ 새해			
4일	5일	6일	7일	8일	9일	10일
11일	12일	13일	14일	15일	16일	17일
18일	19일	20일	21일	22일	23일	24일
25일	26일	27일	28일	29일	30일	31일
2월 1일	2일	3일	4일	5일	6일	7일
8일	9일	10일	11일	12일	13일	14일
15일	16일	17일	18일	19일	20일	21일
★ 설날 연휴	★ 설날	★ 설날 연휴				
22일	23일	24일	25일	26일	27일	28일

Study Schedule

First Semester (March to June) Course: Open-Source SW Analysis (Bigdata)

2026 KCC (July)

In Jeju

강의계획서 [2025년도 1 학기]

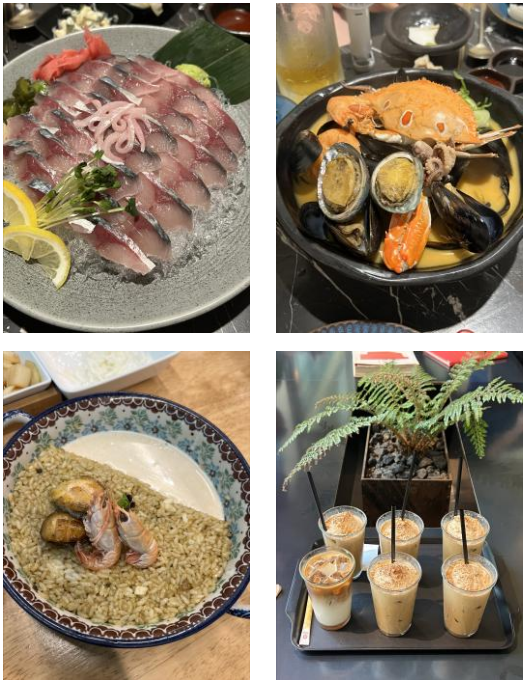
출력

교과목 기본정보(Course Information)

교과목명 Course Title	오픈소스SW분석(빅데이터)	학점 Credits	3
교과목 코드 Course Code	558020-1	이수영역	전공선택
주수강대상	SW융합대학 소프트웨어학과	언어 Language	
강의형태		강의실	목9,10,11,12,13,14(소프트305)
시간구분	이론(3) 실험(0) 실습(0) 실기(0) 설계(0)	사이버강의	웹보조수업
강의유형	대면수업		
원격수업미리보기			

담당교수

담당교수	성명	신호진	직급	강사	최종학위	공학사
	소속	SW융합대학 소프트웨어학과		연구실		
	전화번호	---		e-mail	ghwls03s@gmail.com	
	관심분야					



2026 Winter RocksDB Study Kick-off Meeting

Yongmin Lee, Dayeon Wee

<http://sslslab.dankook.ac.kr/>, <https://sslslab.dankook.ac.kr/~choijm>

Thank You Q & A ?

Presentation by Yongmin Lee
nascarf16@dankook.ac.kr

2026 Winter RocksDB Study

Task for this Study

- ◎ To develop an accurate and in-depth understanding of RocksDB
- ◎ To freely present and discuss individual ideas
- ◎ To formalize these ideas into a research paper

Schedule

- ◎ 1st ~ 2nd week: Theory-based Lecture
- ◎ 3rd ~ 6th week: Presentation-based Lecture
- ◎ 7th week: Final Presentation

오픈 소스 SW 심화 스터디

비정형 빅데이터를 위한 RocksDB 구조 분석과 기술 동향 이해



목표	스터디 상세 정보
구글, 페이스북, 아마존 등에서 사용하는 빅데이터 관리 SW 이해 대표적인 카-벨류 DB인 RocksDB 소스 코드 수준 분석 LSM-tree, 집약, 검색, 불륨 필터 등 핵심 자료구조 이해 실질적이고 전문가 수준의 오픈 소스 분석 및 기여	일정 2026년 1/5일 ~ 2/27일 (매주 1번씩 세미나 진행, 화요일 오후 2-4시 예정) 장소 단국대학교 SW-HCT관 301호 강사 최종무/유시환 교수, 관련 대학원 학생들 세미나 내용 <ul style="list-style-type: none">1주차 빅데이터 특성, 관리 SW 소개, RocksDB 소스 코드 수준 소개2주차 RocksDB 핵심 구조, LSM-tree, 집약, 검색3주차 RocksDB 최근 연구 동향, Wisckey, SILK, Remix 등4주차 RocksDB interface 분석(학생 2-3명 발표, 교수/대학원생 지도)5주차 RocksDB compaction 분석(학생 2-3명 발표, 교수/대학원생 지도)6주차 RocksDB lookup 분석(학생 2-3명 발표, 교수/대학원생 지도)7주차 RocksDB new approach 분석(학생 2-3명 발표, 교수/대학원생 지도)8주차 새로운 기법 구현 및 성능 평가, (가능하면) 논문 작성 위 내용을 진행함에 따라 변경 가능
대상 시스템 프로그래밍, 자료구조 수업 이수자 중 위의 내용에 관심 있는 사람. 쓰레드, 파일, 동기화, 메이지 캐시 학교에서 배운 내용이 등이 실제 산업계 SW에서 어떻게 사용되는지 관심이 있는 사람. 다양한 워크로드에서 실행 및 성능 분석 관심이 있는 사람. C 또는 C++를 다룰 수 있으며 위의 내용에 관심있는 전공자/비전공자 및 재학생/졸업생 누구나 참여가능 지원 연락처 - 이윤민 (nascarf16@dankook.ac.kr) 위다연 (wida10@dankook.ac.kr)	

주관  단국대학교 SW중심대학 OSS 센터  IITP 정보통신기획평가원  StarLab.

2026 Winter RocksDB Study

Task for this Study

- ◎ To develop an accurate and in-depth understanding of RocksDB
- ◎ To freely present and discuss individual ideas
- ◎ To formalize these ideas into a research paper

Schedule

- ◎ 1st ~ 2nd week: Theory-based Lecture
- ◎ 3rd ~ 6th week: Presentation-based Lecture
- ◎ 7th week: Final Presentation

오픈 소스 SW 심화 스터디

비정형 빅데이터를 위한 RocksDB 구조 분석과 기술 동향 이해



목표

구글, 페이스북, 아마존 등에서 사용하는 빅데이터 관리 SW 이해
대표적인 키-밸류 DB인 RocksDB 소스 코드 수준 분석
LSM-tree, 집합, 잠적, 불용 밀터 등 핵심 자료구조 이해
실질적이고 전문가 수준의 오픈 소스 분석 및 기여

대상

시스템 프로그래밍, 자료구조 수업 이수자 중 위의 내용에 관심 있는 사람.
쓰레드, 파일, 동기화, 페이지 캐시 학교에서 배운 내용이 많이 실제 산업체 SW에서 어떻게 사용되는지 관심이 있는 사람.
다양한 워크로드에서 실험 및 성능 분석 관심이 있는 사람.
C 또는 C++를 다룰 수 있으며 위의 내용에 관심있는 전공자/비전공자 및 재학생/졸업생 누구나 참여가능
지원 연락처 : 이윤진 (nascart16@dankook.ac.kr) 위다연 (wida10@dankook.ac.kr)

스터디 상세 정보

일정 2026년 1/5일 ~ 2/27일
(매주 1번씩 세미나 진행, 화요일 오후 2-4시 예정)
장소 단국대학교 SW-ICT관 301호
강사 최종우/유시환 교수, 관련 대학원 학생들
세미나 내용
1주차 빅데이터 특성, 관리 SW 소개, RocksDB 소스 코드 수준 소개
2주차 RocksDB 핵심 구조, LSM-tree, 집합, 잠적
3주차 RocksDB 최근 연구 동향, Wickey, SILK, Remix 등
4주차 RocksDB interface 분석(학생 2-3명 발표, 교수/대학원생 지도)
5주차 RocksDB compaction 분석(학생 2-3명 발표, 교수/대학원생 지도)
6주차 RocksDB lookup 분석(학생 2-3명 발표, 교수/대학원생 지도)
7주차 RocksDB new approach 분석(학생 2-3명 발표, 교수/대학원생 지도)
8주차 새로운 기법 구현 및 성능 평가, (가능하면) 논문 작성
위 내용을 진행함에 따라 변경 가능