

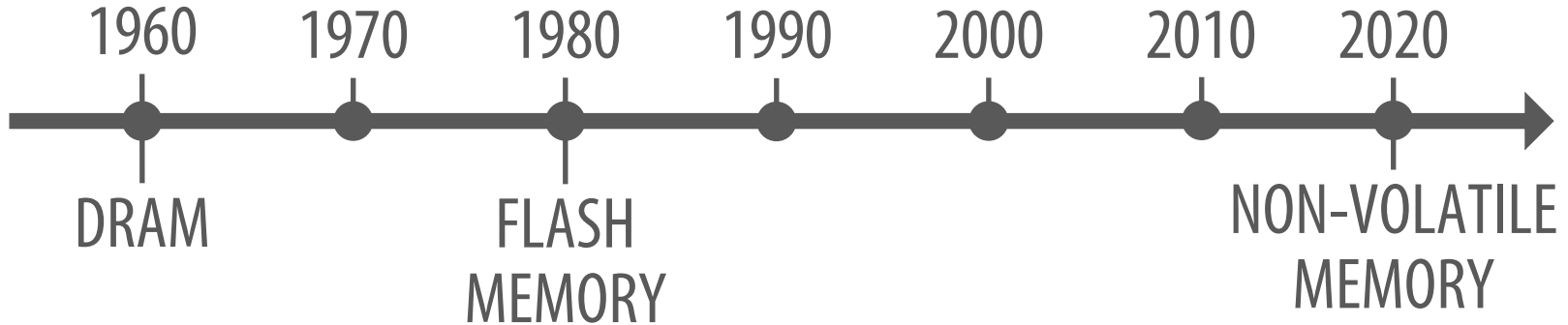
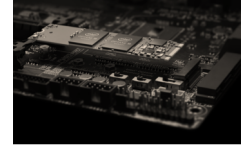
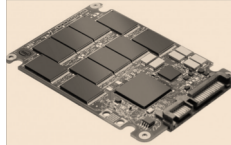
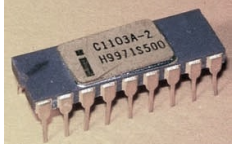
DATA MANAGEMENT ON NON-VOLATILE MEMORY

JOY ARULRAJ

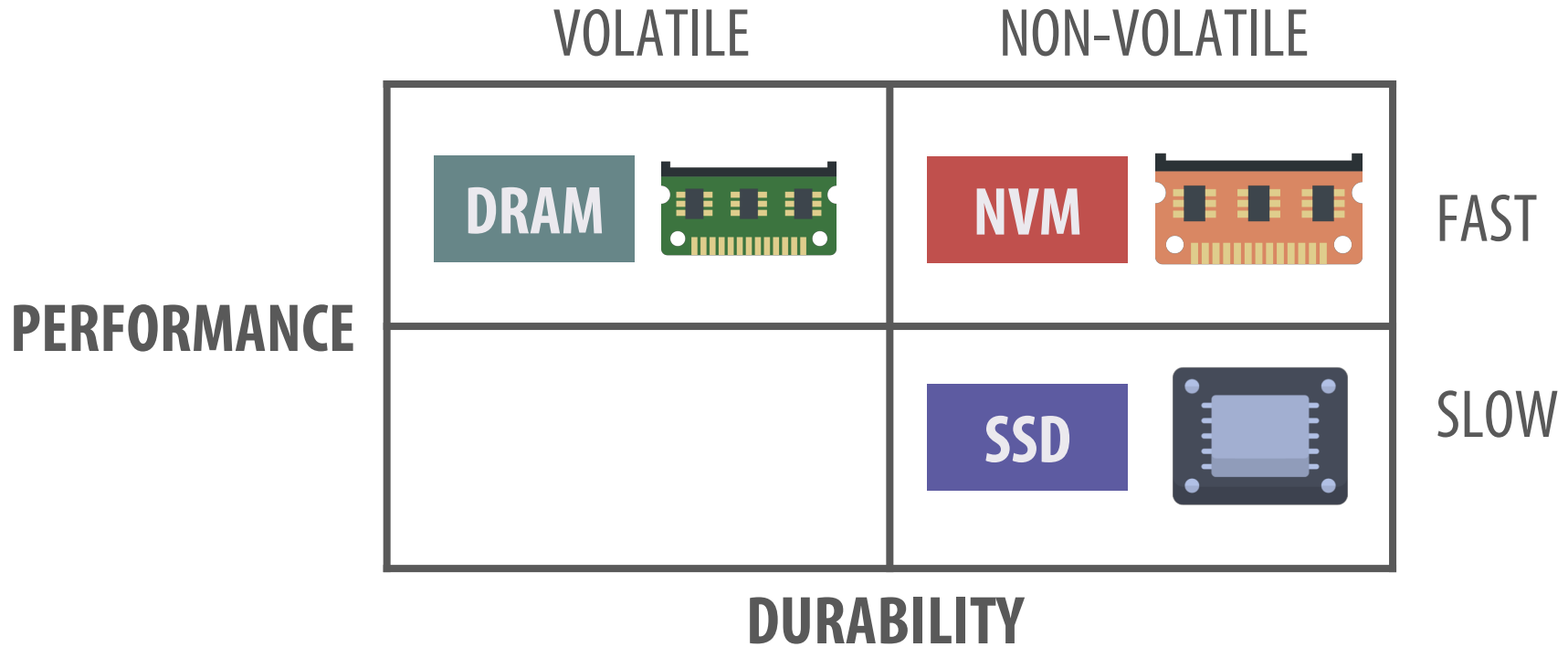
Georgia Institute of Technology

2019 ACM SIGMOD Jim Gray Award Talk

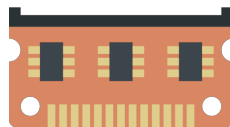
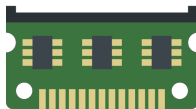
EVOLUTION OF MEMORY TECHNOLOGY



NON-VOLATILE MEMORY [NVM]



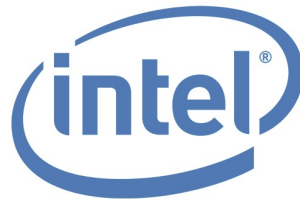
DEVICE CHARACTERISTICS



CHARACTERISTIC	DRAM	NVM	SSD
Device Latency	1x	~10x	1000x
Fine-Grained Access	✓	✓	✗
Durability	✗	✓	✓
High Capacity	✗	✓	✓
Cost/GB	100x	~10x	1x

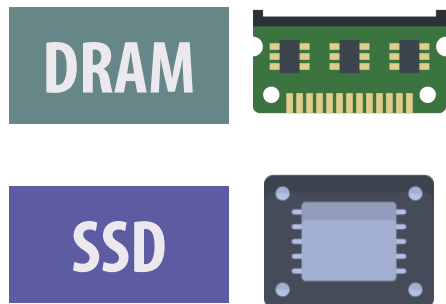
RECENT NVM-RELATED DEVELOPMENTS

- #1: Industry standards
 - *Design specifications*
- #2: Operating system support
 - *Linux 4.8+, Windows 10*
- #3: Architectural support
 - *New assembly instructions*
- #4: Intel has started shipping NVM DIMMs
 - *Optane DC DIMMs (April 2019)*

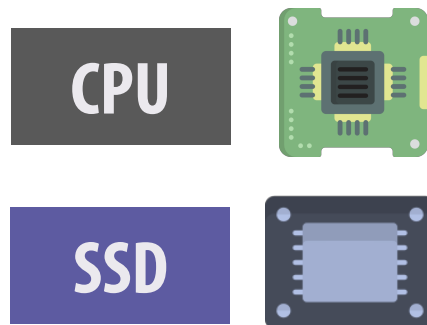


50 YEARS OF DATABASE SYSTEMS RESEARCH

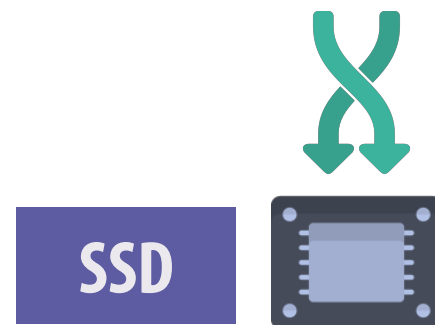
1 SEPARATE MEMORY & STORAGE



2 COMPUTE/STORAGE BALANCE



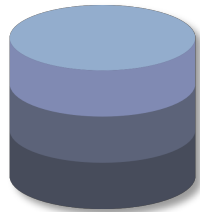
3 RANDOM VS. SEQUENTIAL



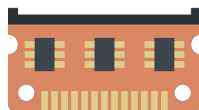
THESIS OVERVIEW

- How to manage data on NVM?
 - *Challenging because of NVM's unique characteristics*

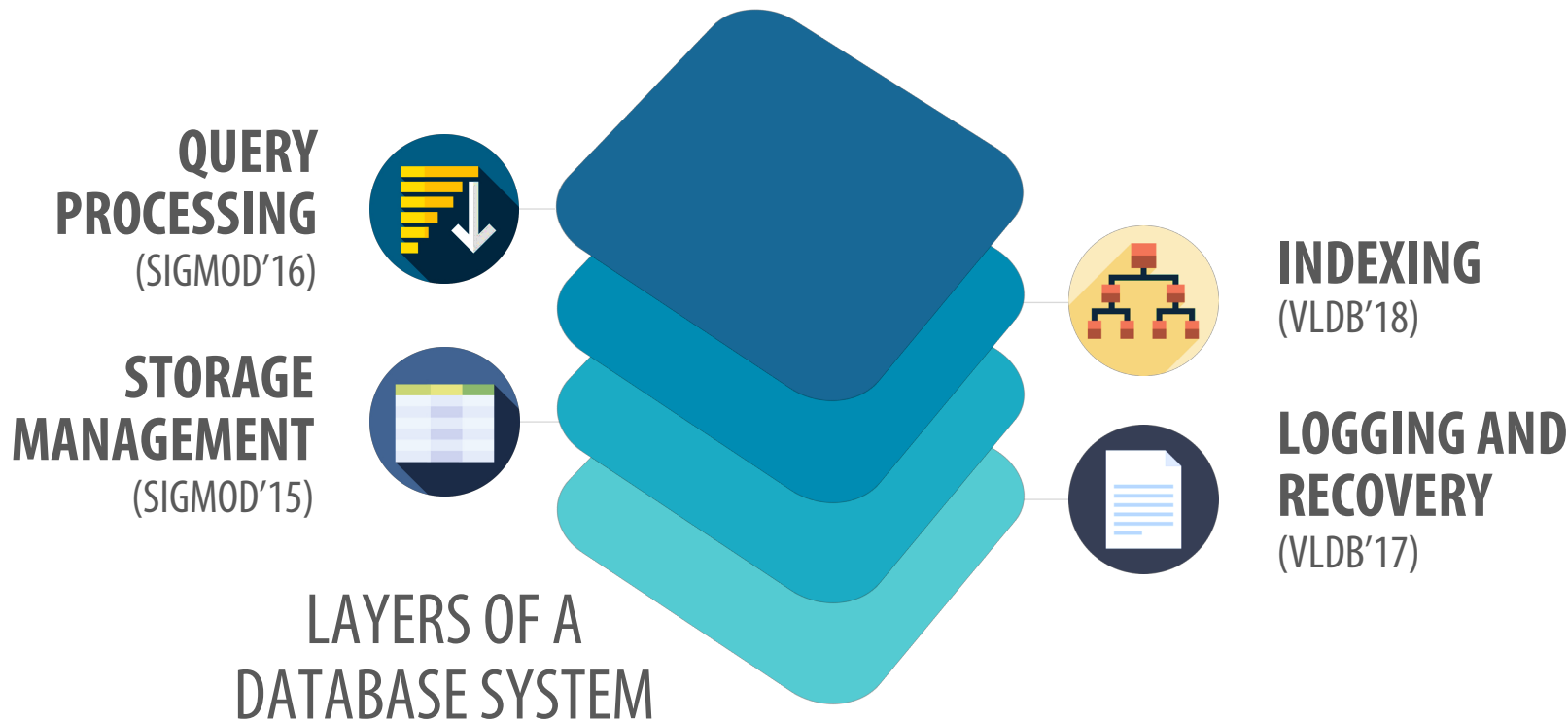
**DATABASE
SYSTEM**



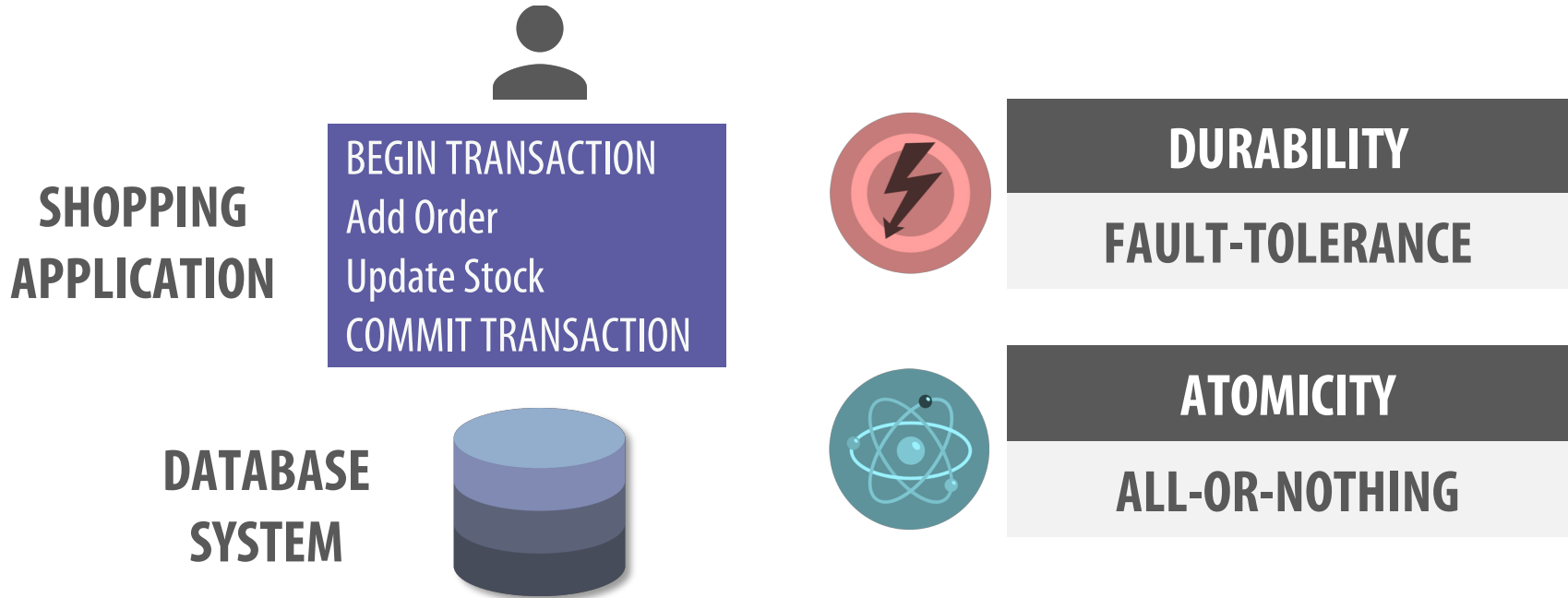
NVM



PELTON NVM DATABASE SYSTEM



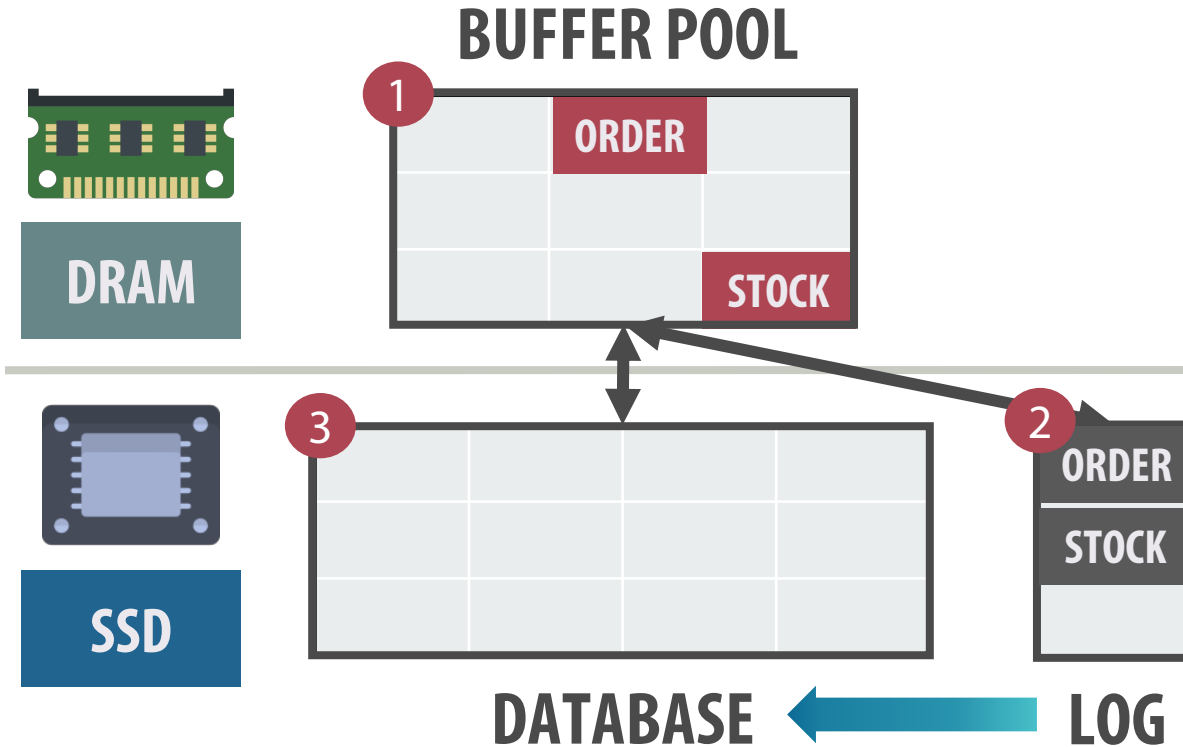
LOGGING & RECOVERY: MOTIVATION



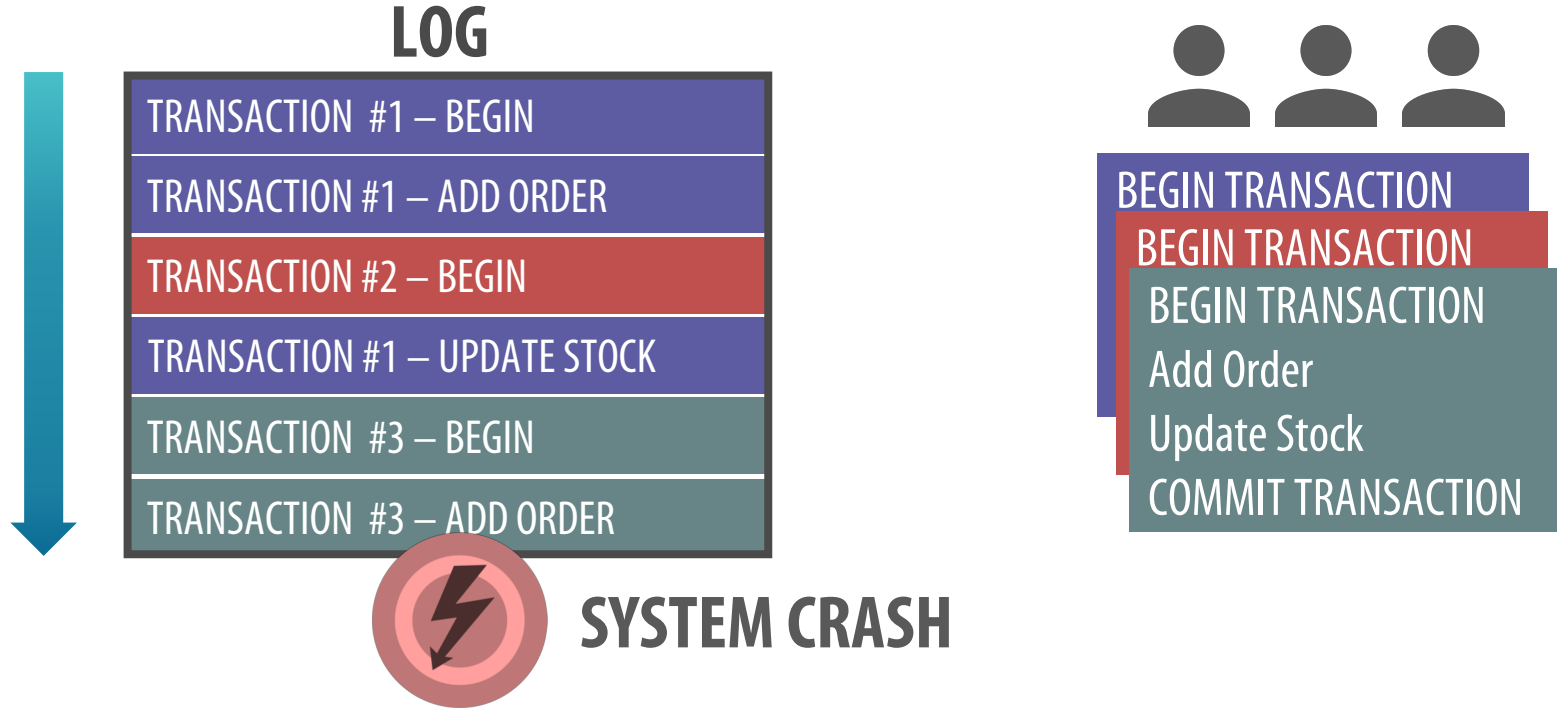
WRITE-AHEAD LOGGING: DURABILITY



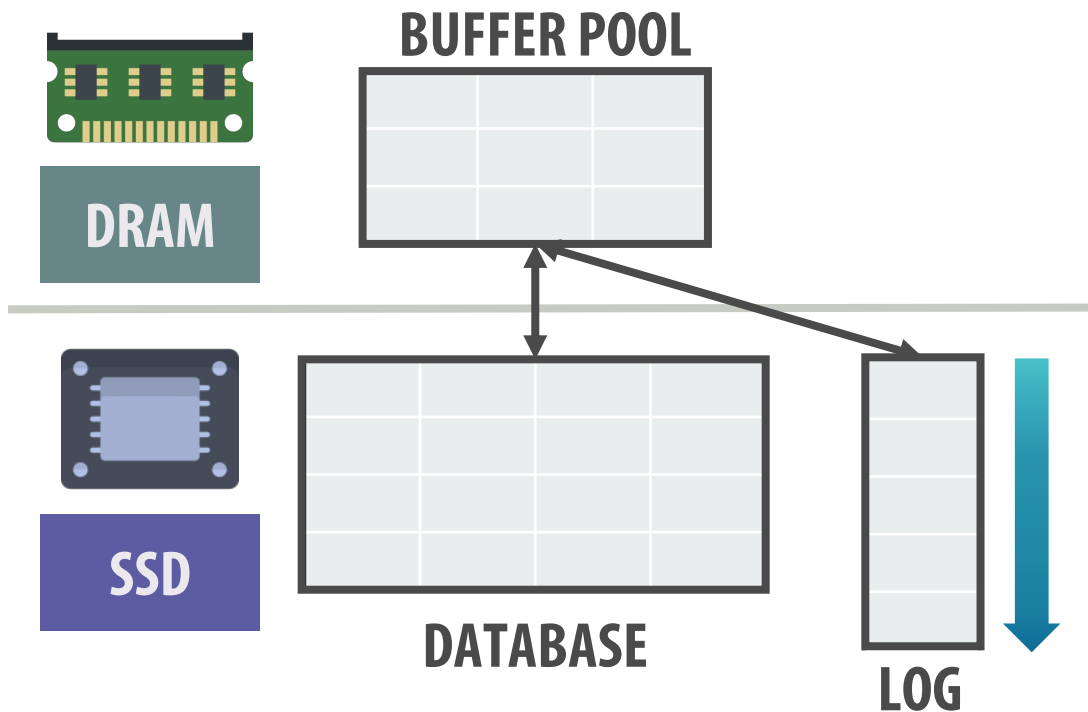
BEGIN TRANSACTION
Add Order
Update Stock
COMMIT TRANSACTION



WRITE-AHEAD LOGGING: ATOMICITY



WRITE-AHEAD LOGGING: RECOVERY PROTOCOL



3 UNDO CHANGES

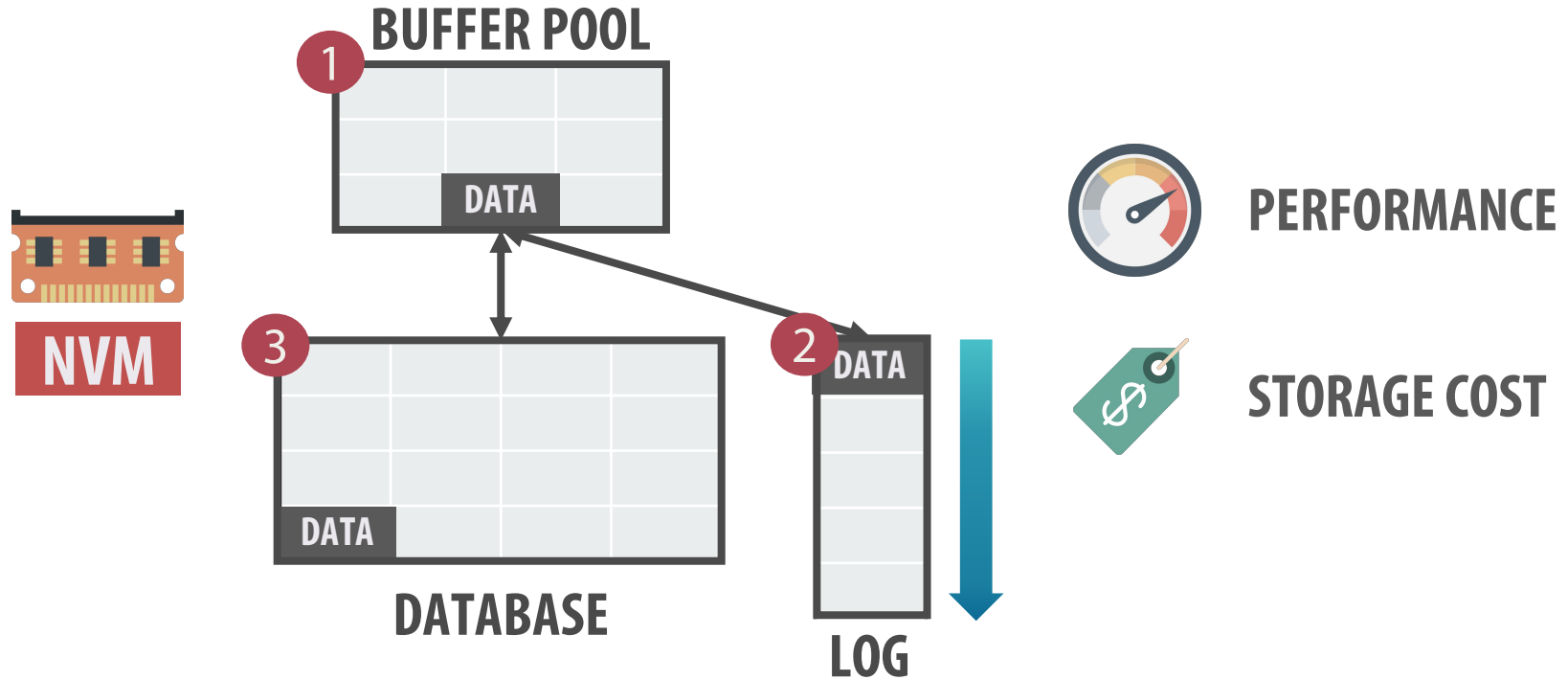
ACTIVE TRANSACTION TABLE

TXN ID	STATUS	LATEST CHANGE
TXN #2	RUNNING	LOG RECORD #7
TXN #3	RUNNING	---

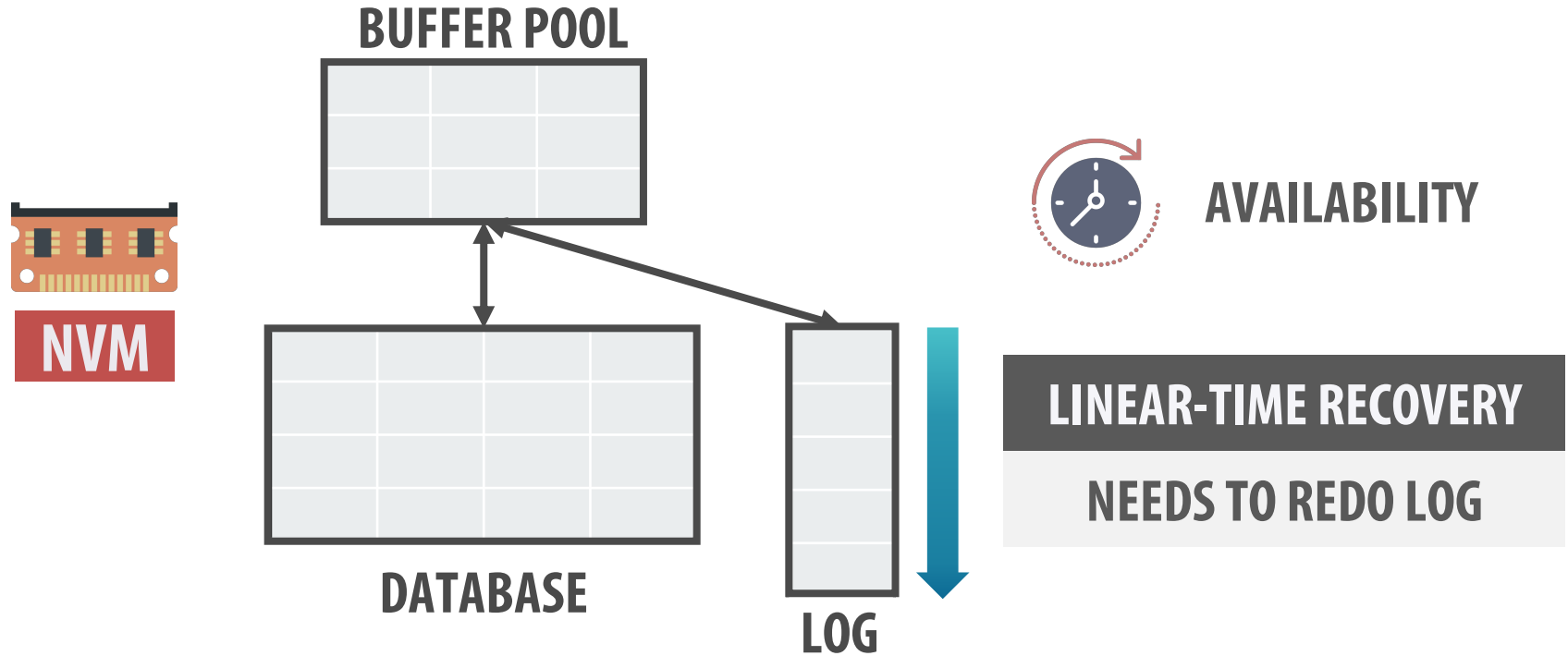
DIRTY PAGE TABLE

PAGE ID	CHANGE THAT DIRTIED PAGE
PAGE #30	LOG RECORD #5
PAGE #40	LOG RECORD #7

PROBLEM #1: DATA DUPLICATION



PROBLEM #2: SLOW RECOVERY



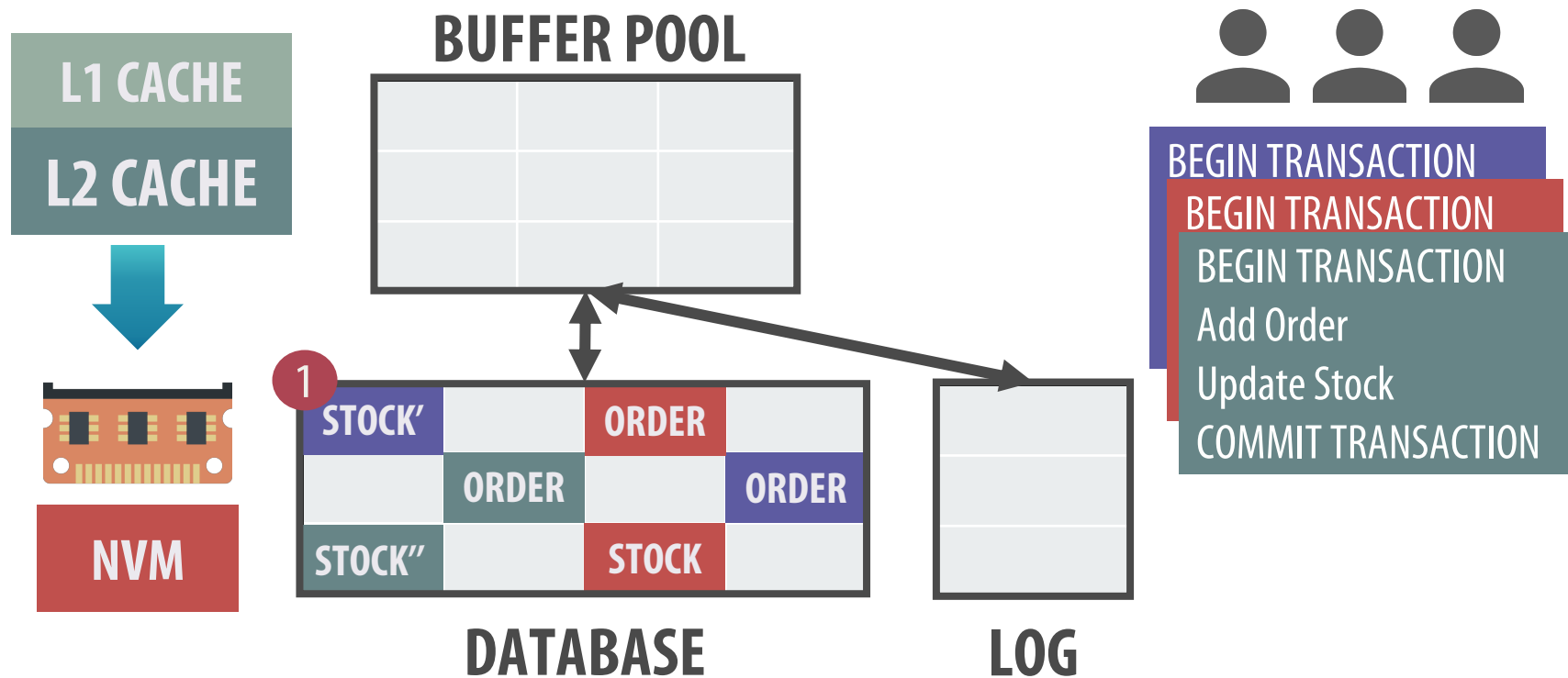
HOW TO IMPROVE PERFORMANCE AND AVAILABILITY ON NON-VOLATILE MEMORY?



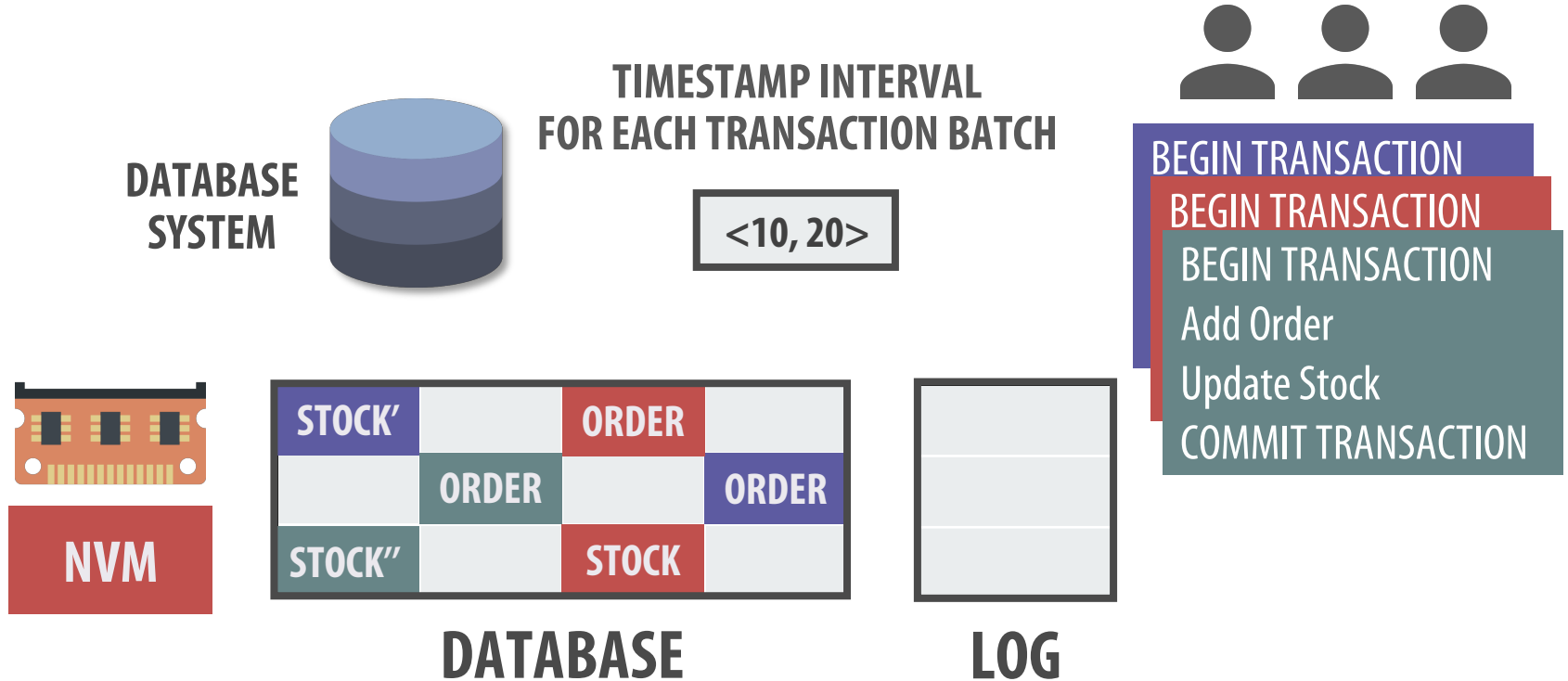
WRITE-BEHIND LOGGING: OVERVIEW

- NVM-centric design
 - *Improves availability by enabling instant recovery*
 - *Provides same guarantees as write-ahead logging*
- Key techniques
 - *Directly propagate changes to the database*
 - *Only record meta-data in log*

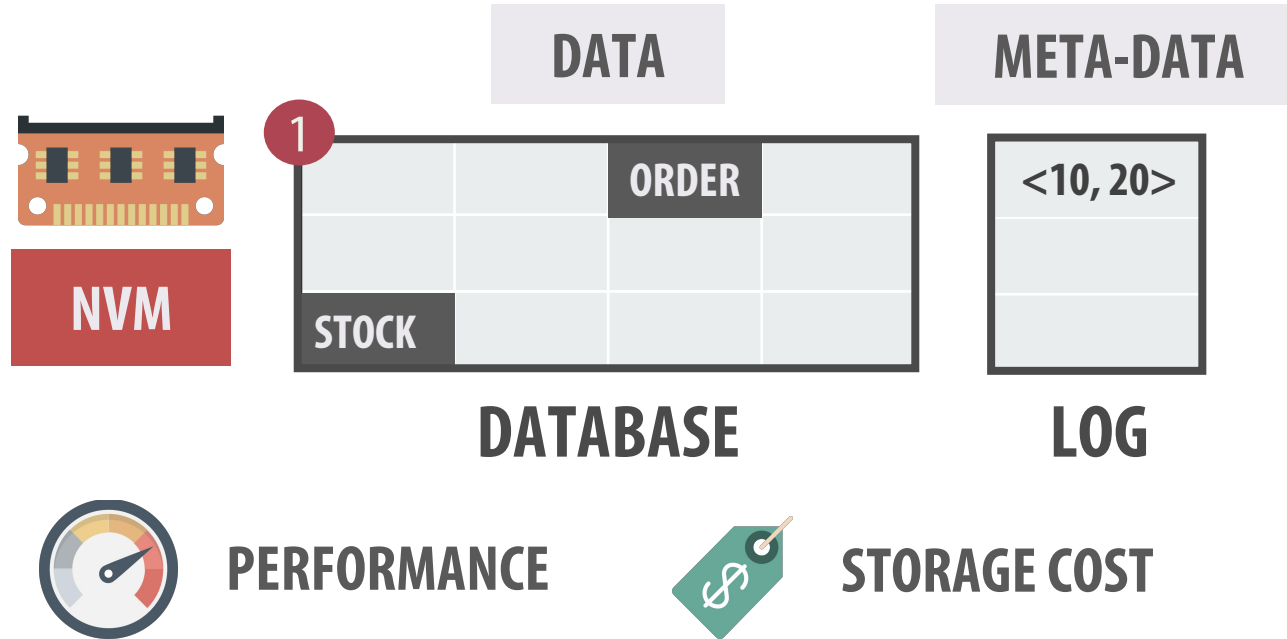
WRITE-BEHIND LOGGING: DURABILITY



WRITE-BEHIND LOGGING: ATOMICITY



SOLUTION #1: NO DATA DUPLICATION



SOLUTION #2: INSTANT RECOVERY

WRITE-AHEAD LOGGING

Linear-Time Recovery

1

ANALYSIS

2

REDO

3

UNDO

WRITE-BEHIND LOGGING

Constant-Time Recovery

1

ANALYSIS



AVAILABILITY

WRITE-BEHIND LOGGING: SUMMARY



AVAILABILITY



PERFORMANCE

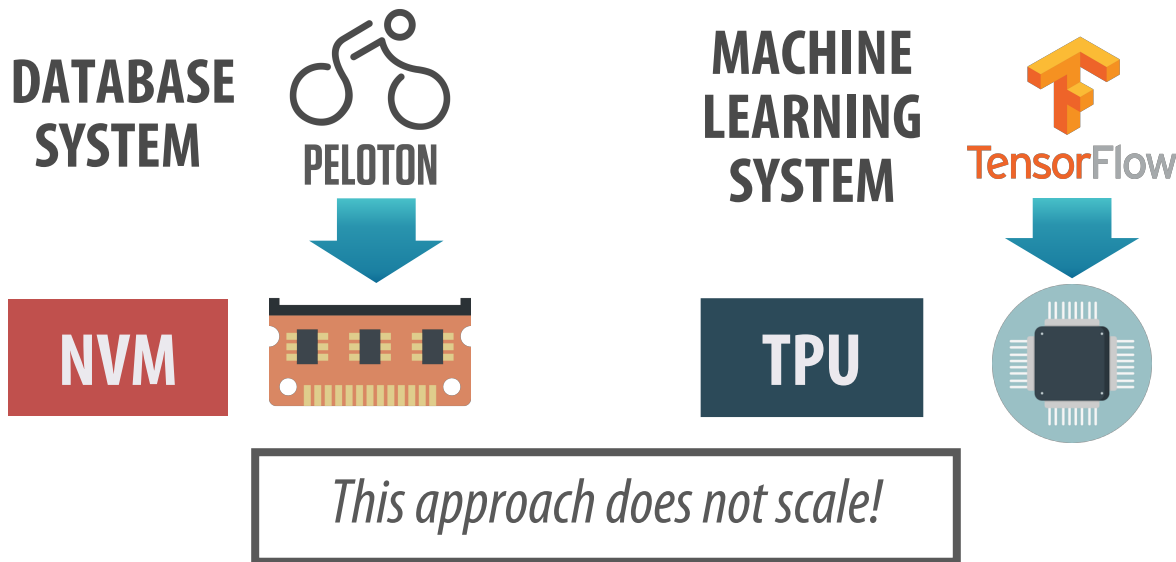


STORAGE COST

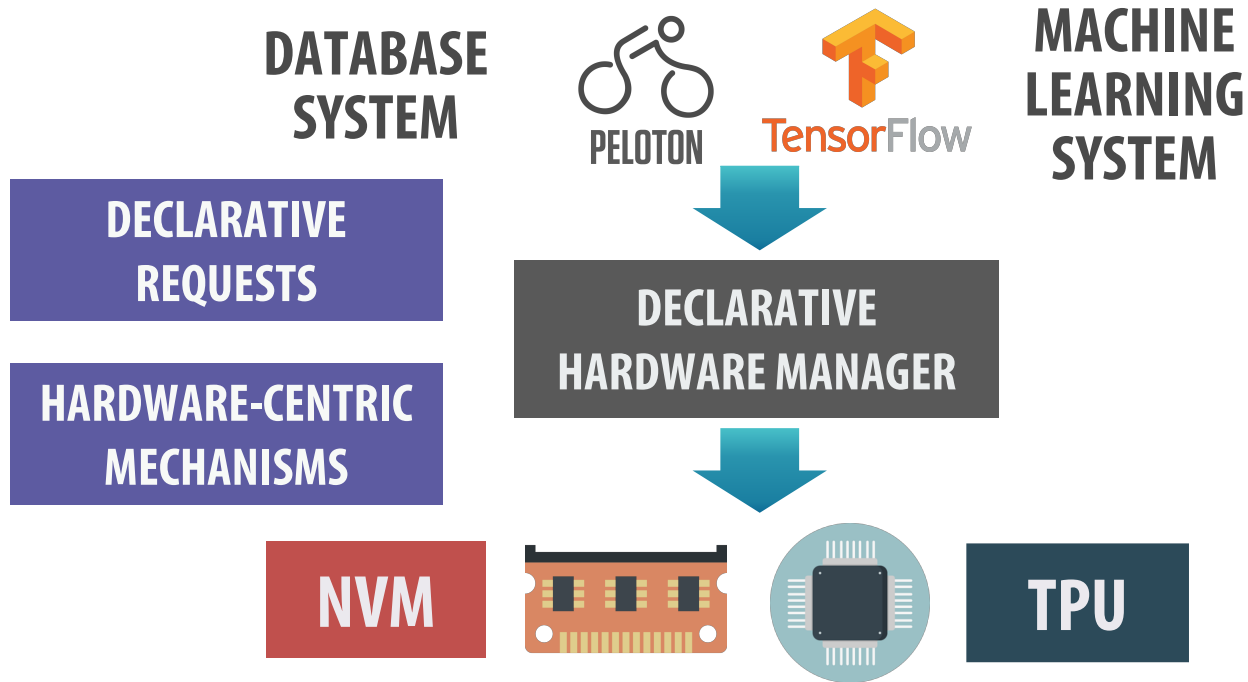
Advances the state of the art by shifting the complexity class of the recovery protocol on NVM

FUTURE RESEARCH

- Most hardware-centric optimizations are system-specific
 - *One-to-one mapping between hardware-centric optimizations & systems*



DECLARATIVE HARDWARE MANAGEMENT



CONCLUSION

- Advent of NVM has begun a new era in systems research
 - *Presented the design of an NVM-centric logging and recovery protocol*
 - *Broader impact on all types of data processing systems*
- This thesis was greatly influenced by several NVM researchers
 - *Ismail Oukid and Wolfgang Lehner, TU Dresden*
 - *Tianzheng Wang and Ryan Johnson, University of Toronto*
 - *Alexander van Renen, Viktor Leis, and Thomas Neumann, TU Munich*
 - *Joel Coburn and Steve Swanson, University of California – San Diego*

ADVISOR



ANDY PAVLO

MENTORS



SAM MADDEN



DONALD KOSSMANN



H.V. JAGADISH

COLLEAGUES & COLLABORATORS

- Subramanya Dulloor, Viktor Leis, Justin Levandoski, Lin Ma, Krishna Teja Malladi, Prashanth Menon, Umar Farooq Minhas, Per-Ake Larson , Matthew Perron, Kai Ren, Jungmin Seo, Vivek Seshadri, Anthony Tomasic, Dana Van Aken, Yingjun Wu, and Michael Zhang.
- This work was done in collaboration with the Intel Science and Technology Center for Big Data, Microsoft Research, and Samsung Research.

END

@joy_arulraj