



英特尔® 傲腾™数据中心级 持久内存存在Elasticsearch 云的应用与实践

侯杰

2019/11, Software Apps Engineer, Intel

议程

英特尔® 傲腾™数据中心级持久内存Elasticsearch云的应用与实践

- 1 英特尔® 傲腾™数据中心级持久内存介绍
- 2 ElasticSearch 内存利用模式
- 3 ElasticSearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化
- 4 ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化
- 5 在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

什么是持久化内存

What is Persistent Memory?

- Byte Addressable
- Cache Coherent
- Memory-like Performance

Why Does it Matter Now?

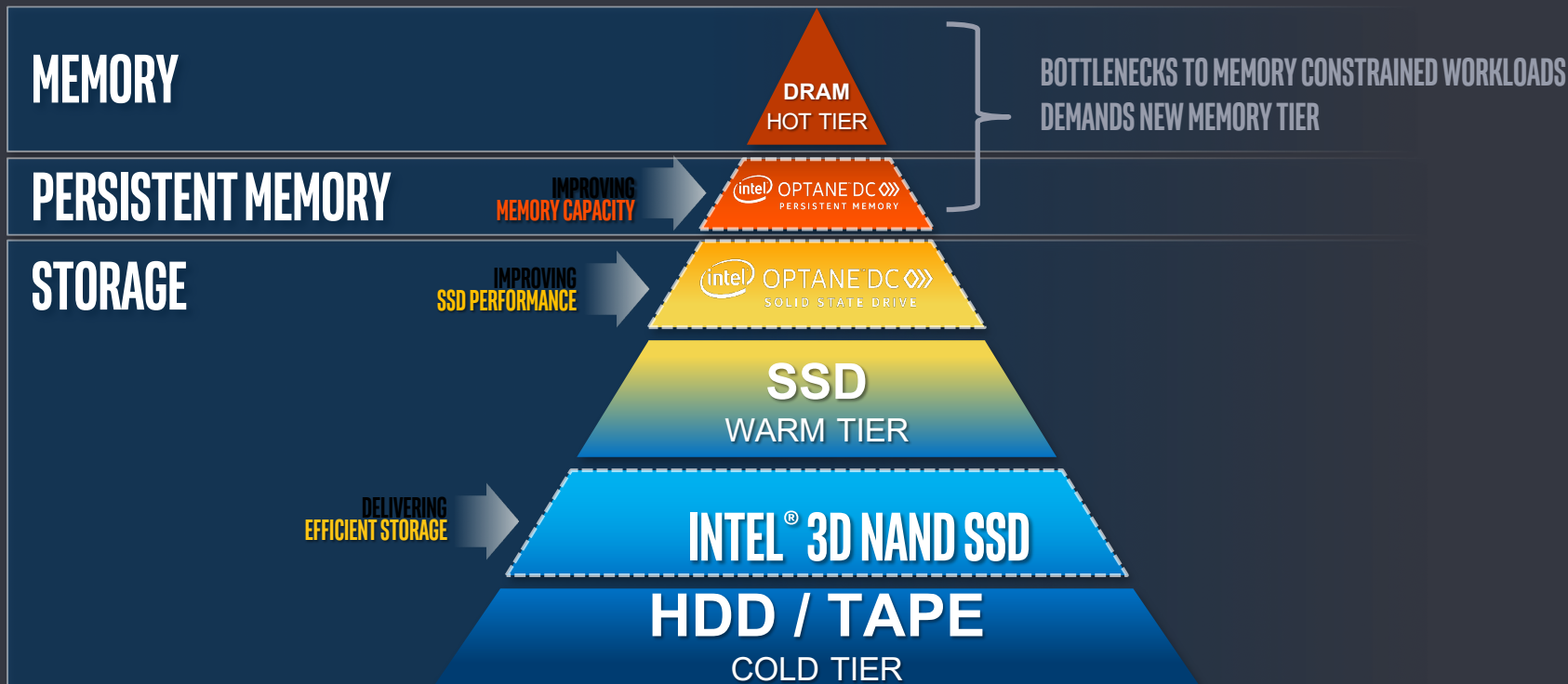
- Adds a new tier between DRAM and Block Storage
- Larger Capacity, Higher Endurance, Consistent low latency
- Ability to do in-place persistence

What's the Impact on the Applications?

- Need Ways to Enable Access to New High Performance Tier
- May need to re-architect to unlock the new features and performance

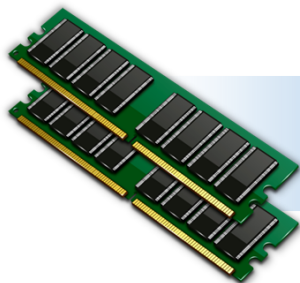
英特尔® 傲腾™数据中心级持久内存介绍

存储架构的重构

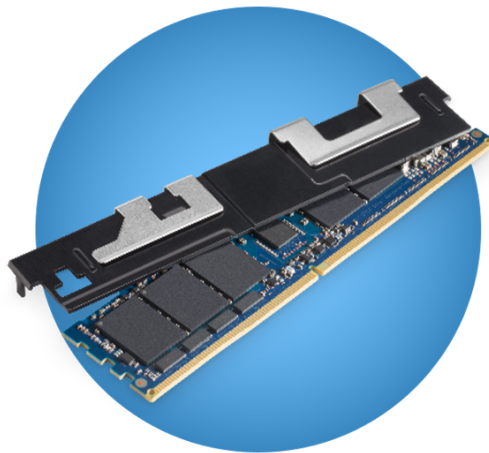


英特尔® 傲腾™数据中心级持久内存介绍

硬件产品概况



FAST MEMORY



**ENHANCE DATA INSIGHTS BY REDEFINING
THE MEMORY & STORAGE HIERARCHY**

intel OPTANE™ DC
PERSISTENT MEMORY



**SIZE AND DATA PERSISTENCE
OF STORAGE**

Supported on future
Intel® Xeon® Scalable Processors
Platinum and Gold SKUs



英特尔® 傲腾™数据中心级持久内存介绍

工作模式介绍

App Direct Mode

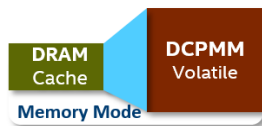
- Fast Recovery from power loss, RDMA for persistent data replication
- Byte Addressable Memory, Apps manage exposed PMEM regions via DAX calls
- OS sees DRAM + AD as total Memory capacity
- OS/VMM Requires PMEM aware file system

Memory Mode

- Higher system capacity vs DRAM and/or lower TCO
- No App changes required
- Encryption key is trashed & volatile data cannot be logically located/accessed after power cycle
- OS sees MM as total Memory capacity, not DRAM; DRAM as direct-mapped Cache

Volatile Memory

DRAM caching



Persistent Memory

No caching

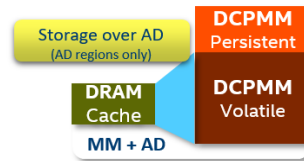


Mixed Mode

- Simultaneously have high capacity and/or lower TCO memory while reserving a portion of the DCPMM for persistent memory usage.

Volatile & Persistent

Memory in separate Regions



议程

英特尔® 傲腾™数据中心级持久内存Elasticsearch云的应用与实践

- 1 英特尔® 傲腾™数据中心级持久内存介绍
- 2 Elasticsearch 内存利用模式**
- 3 Elasticsearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化
- 4 Elasticsearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化
- 5 在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

ElasticSearch 内存利用模式

ElasticSearch memory usage model

Java heap memory

Less than 32GB

Off heap memory

Used by Lucene index

Elasticsearch typical deployment configuration for memory

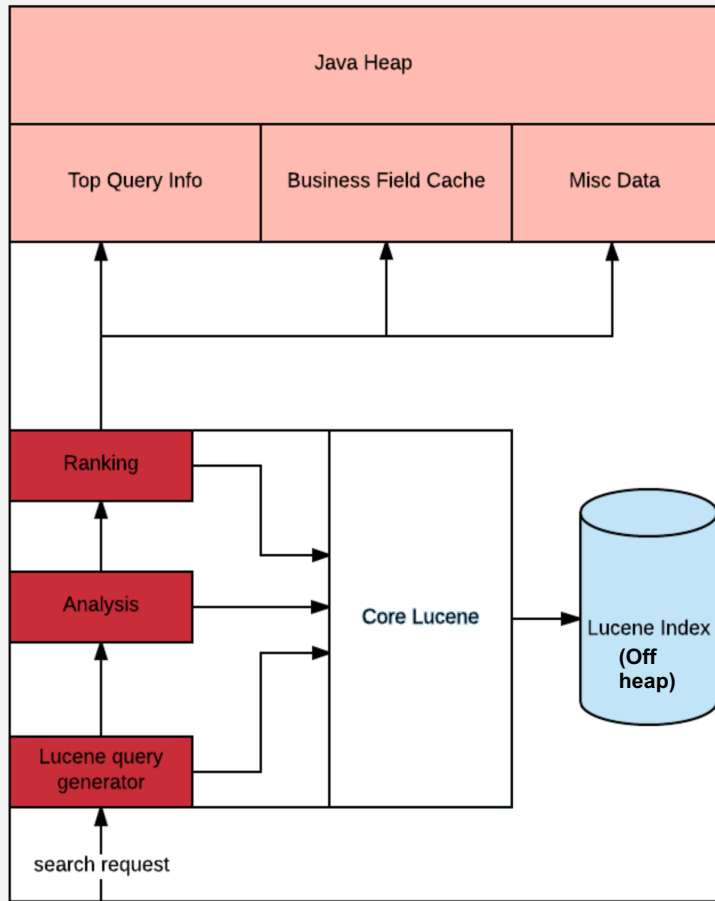
jvm.options

-Xms and -Xmx : for heap memory size.

elasticsearch.yml

index.store.preload:

make off-heap memory larger.



https://www.elastic.co/guide/en/elasticsearch/guide/current/heap-sizing.html#_give_half_your_memory_to_lucene

议程

英特尔® 傲腾™数据中心级持久内存Elasticsearch云的应用与实践

- 1 英特尔® 傲腾™数据中心级持久内存介绍
- 2 ElasticSearch 内存利用模式
- 3 ElasticSearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化**
- 4 ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化
- 5 在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

ElasticSearch Java堆内存利用英特尔傲腾数据中心级持久内存的优化

Memory Mode

Large Memory Capacity

No Software/Application Changes

Data is Volatile

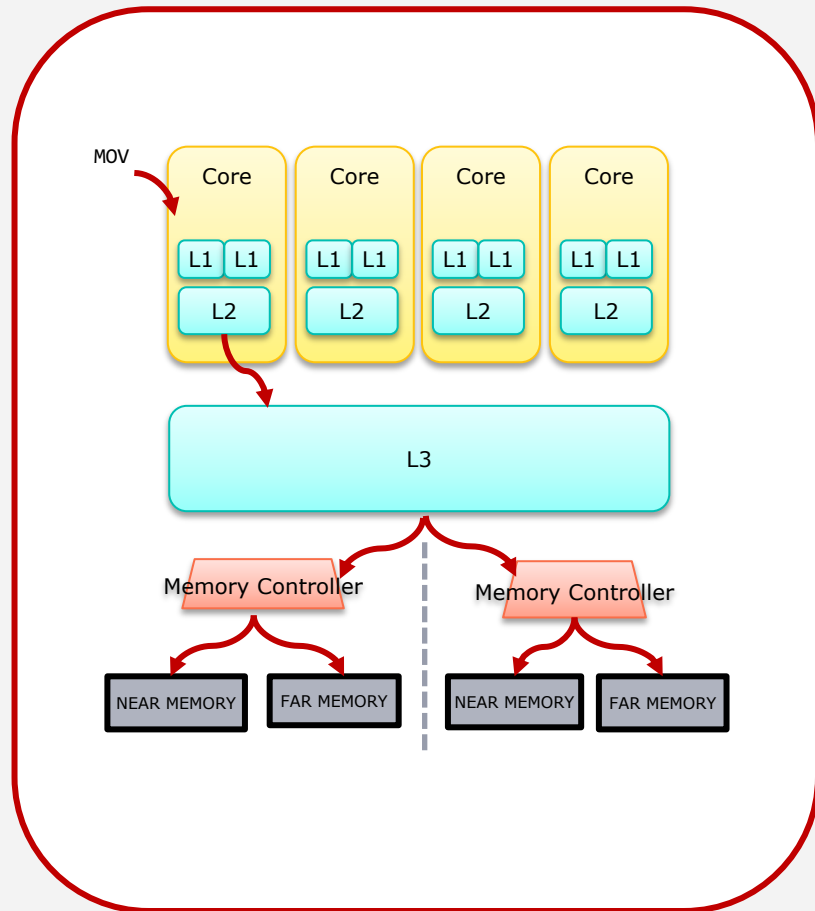
- Volatile mode key cleared and regenerated every power cycle

DRAM As WB Cache

- Managed by Memory Controller
- Within the same memory controller and not across

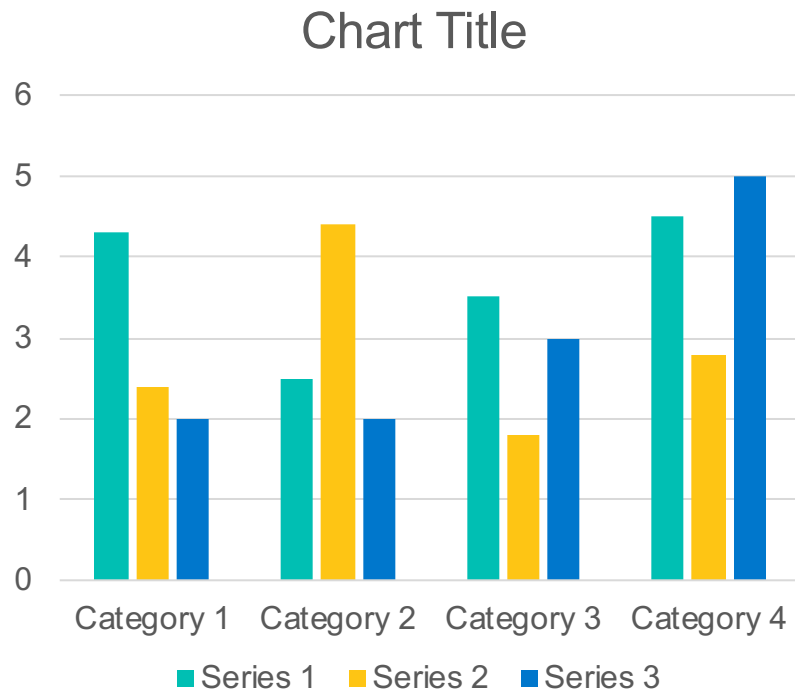
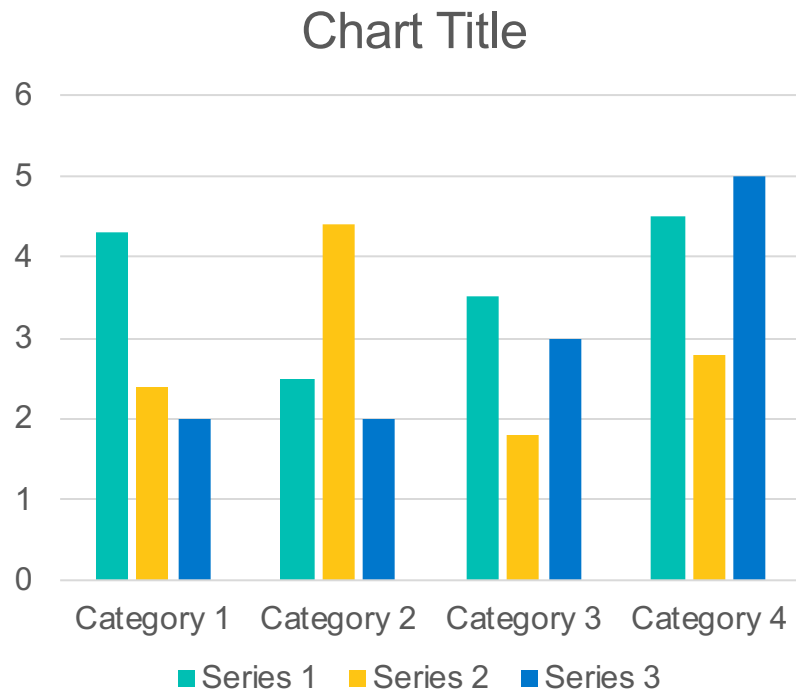
Latency

- Same as DRAM for Cache hit
- DCPMM+DRAM for a Miss



ElasticSearch Java堆内存利用英特尔傲腾数据中心级持久内存的优化

Memory Mode 性能数据



ElasticSearch Java堆内存利用英特尔傲腾数据中心级持久内存的优化

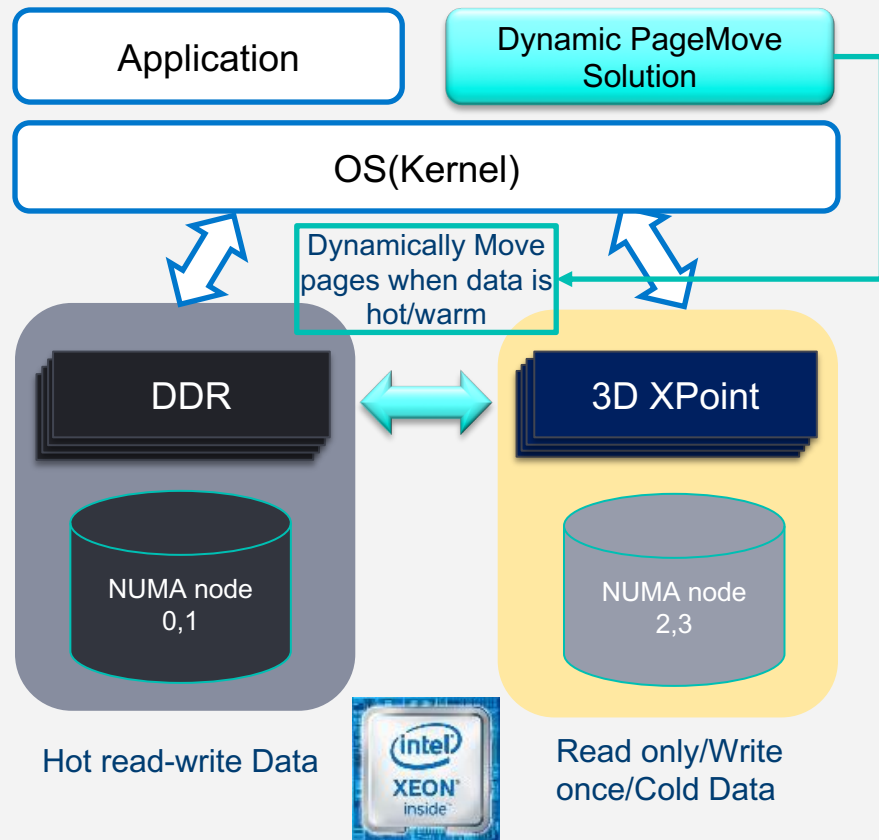
Dynamic PageMove Solution

Advantage:

- Capacity = AEP + DDR
- Predicable performance – when and which data move between AEP and DDR controlled by user

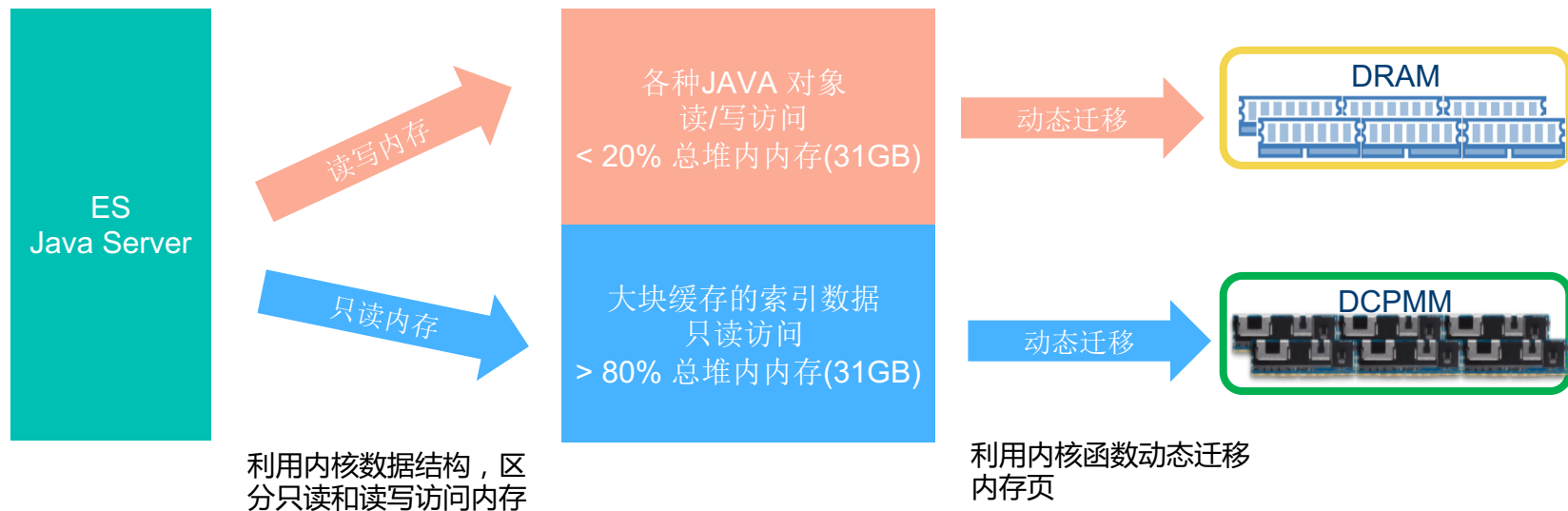
Opportunities:

- No kernel patch needed
- All done in user space
- Target applications change nothing



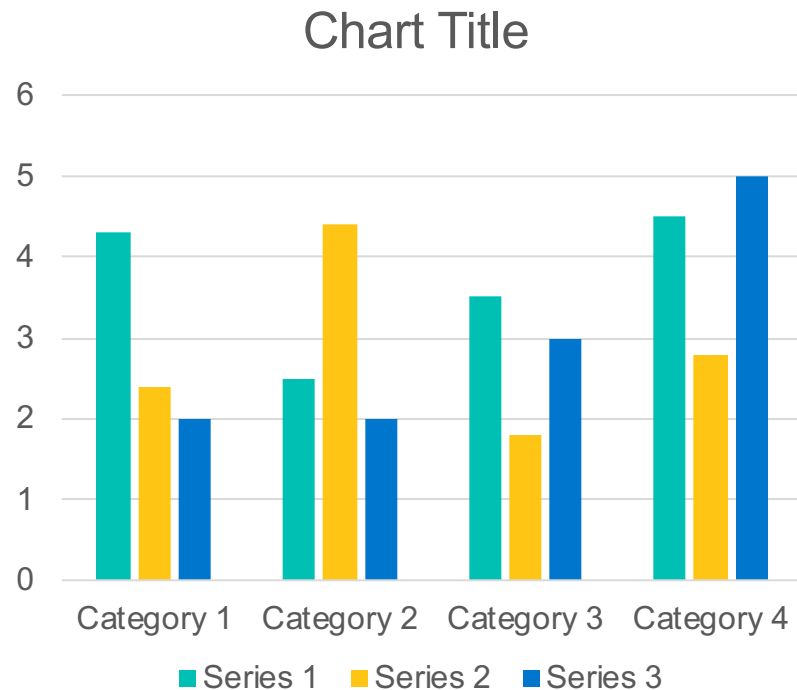
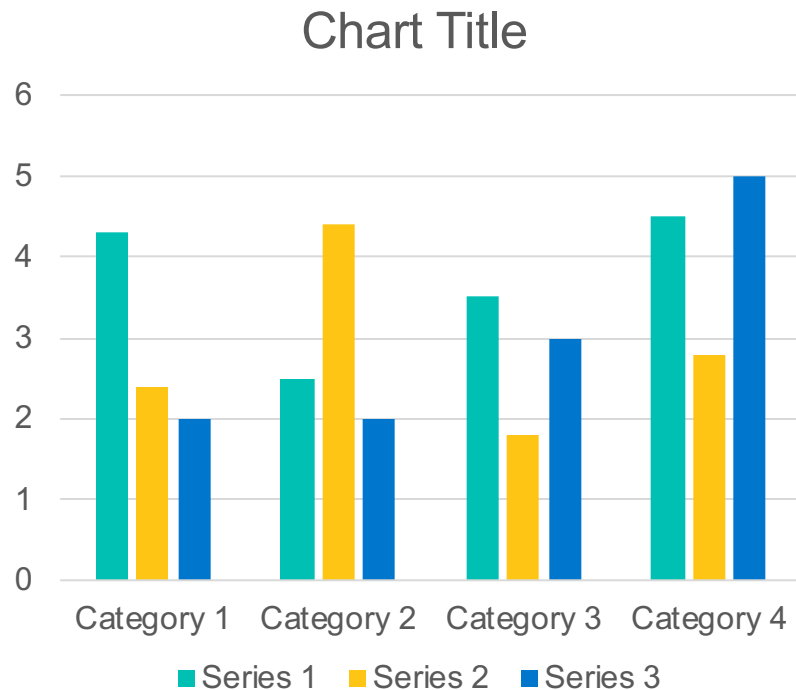
ElasticSearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化

Dynamic PageMove Solution



ElasticSearch Java堆内存利用英特尔傲腾数据中心级持久内存的优化

Dynamic PageMove Solution 性能数据



议程

英特尔® 傲腾™数据中心级持久内存Elasticsearch云的应用与实践

- 1 英特尔® 傲腾™数据中心级持久内存介绍
- 2 ElasticSearch 内存利用模式
- 3 ElasticSearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化
- 4 ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化**
- 5 在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化

APP Direct Mode

Operates in Blocks like SSD/HDD

- Traditional read/write
- Works with Existing File Systems
- Atomicity at block level
- Block size configurable

NVDIMM Driver required

- Support starting Kernel 4.2

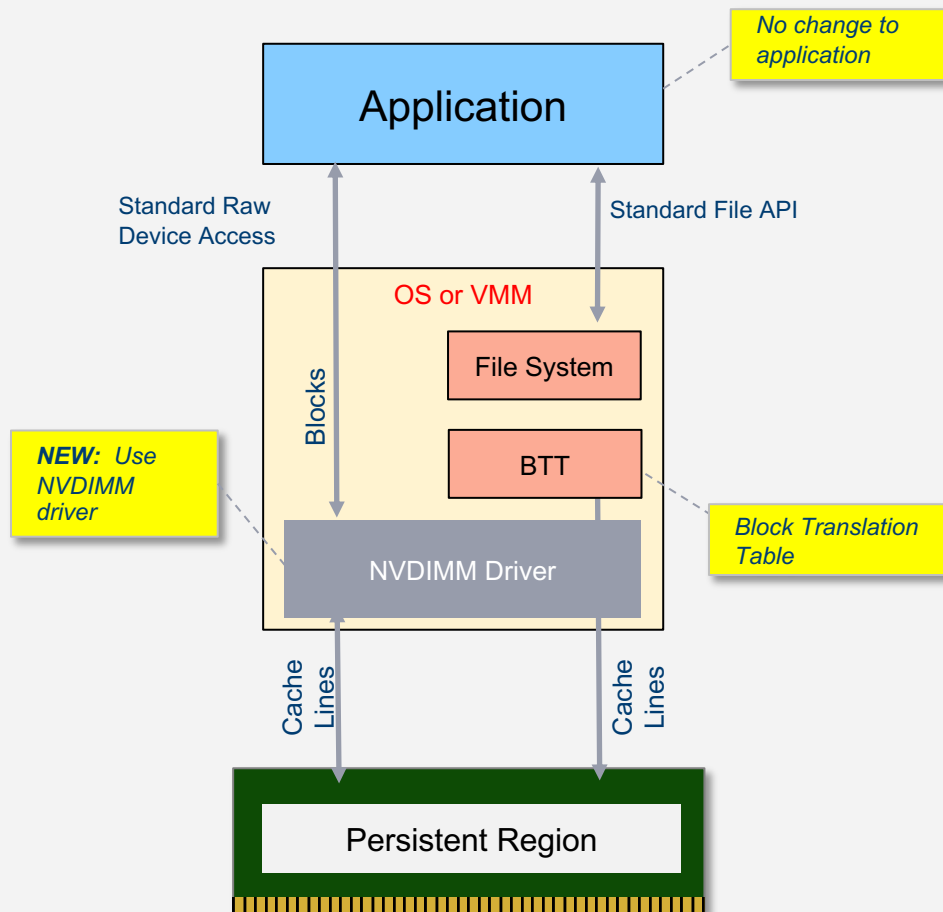
Scalable Capacity

- Up to 3TB/Socket

High Performance Block Storage

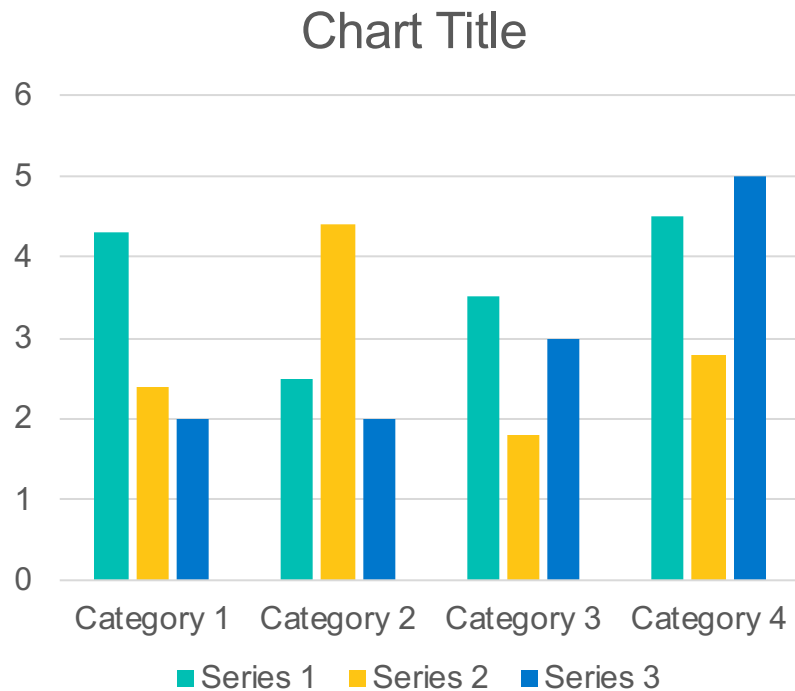
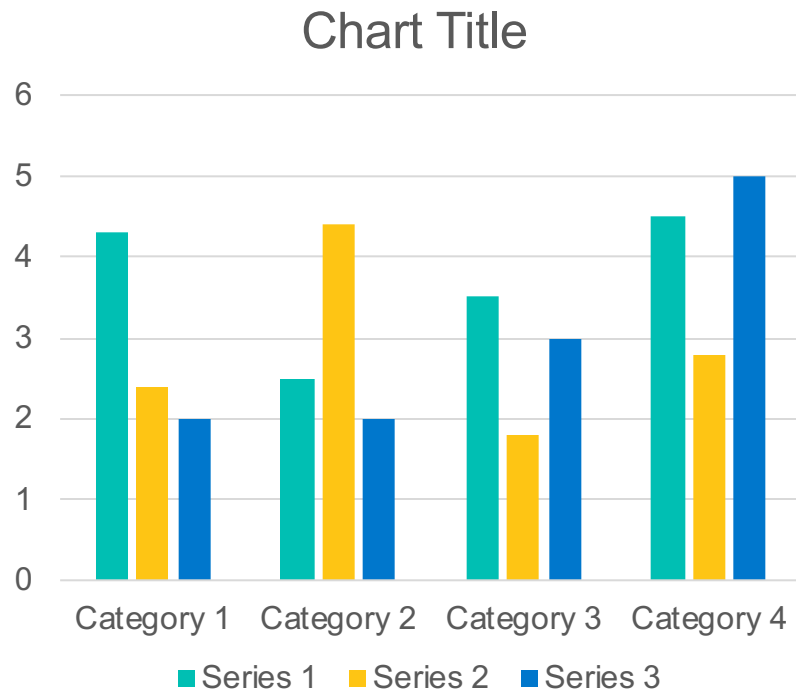
- Low Latency, higher BW, High IOPs

Storage over App Direct



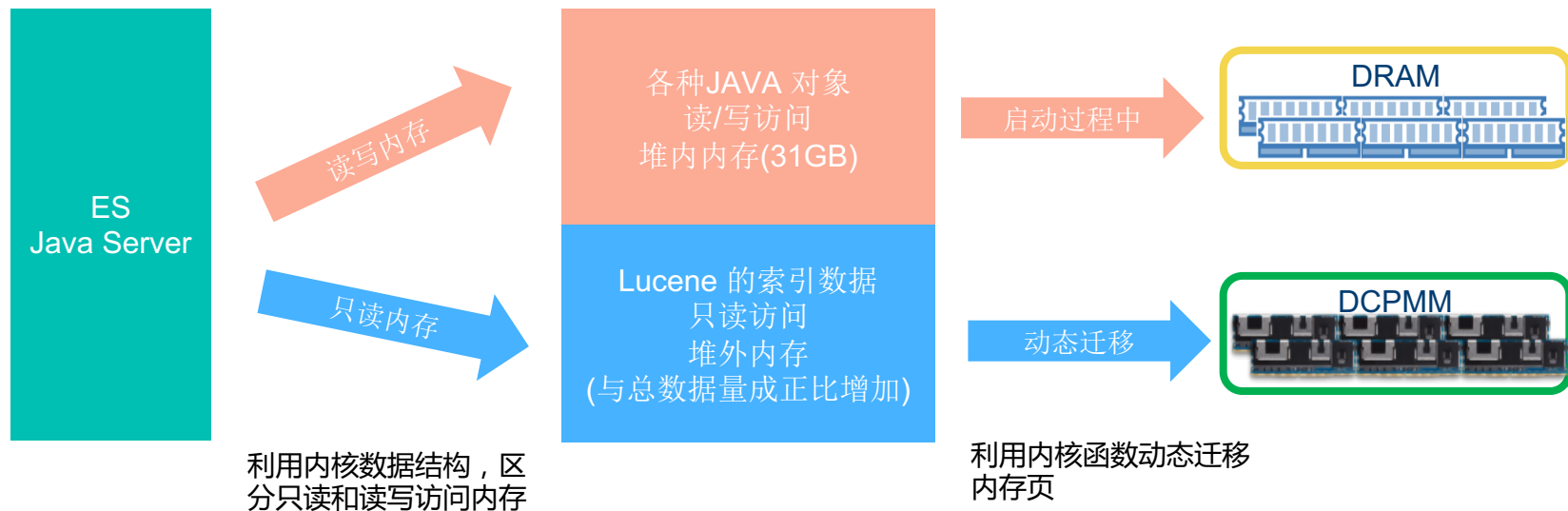
ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化

APP Direct Mode 性能数据



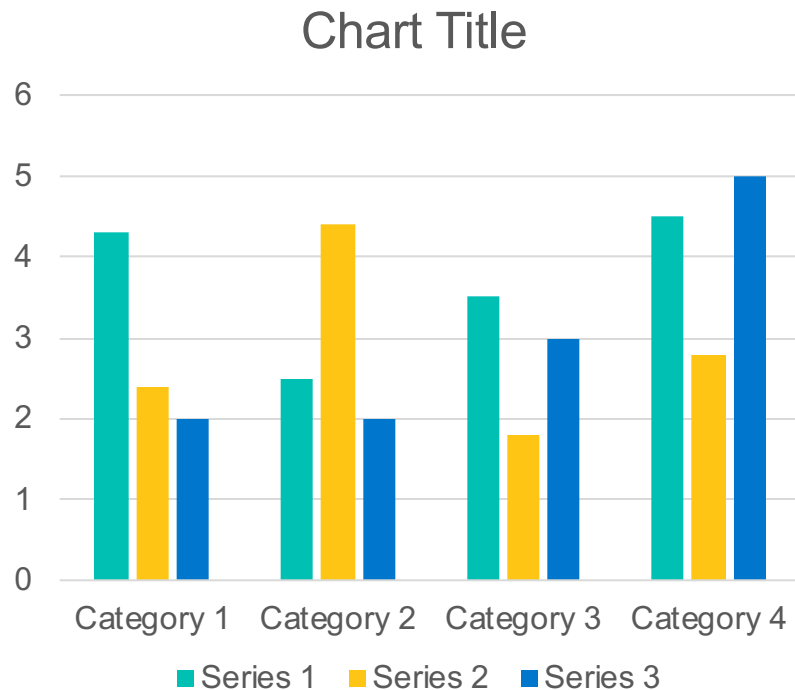
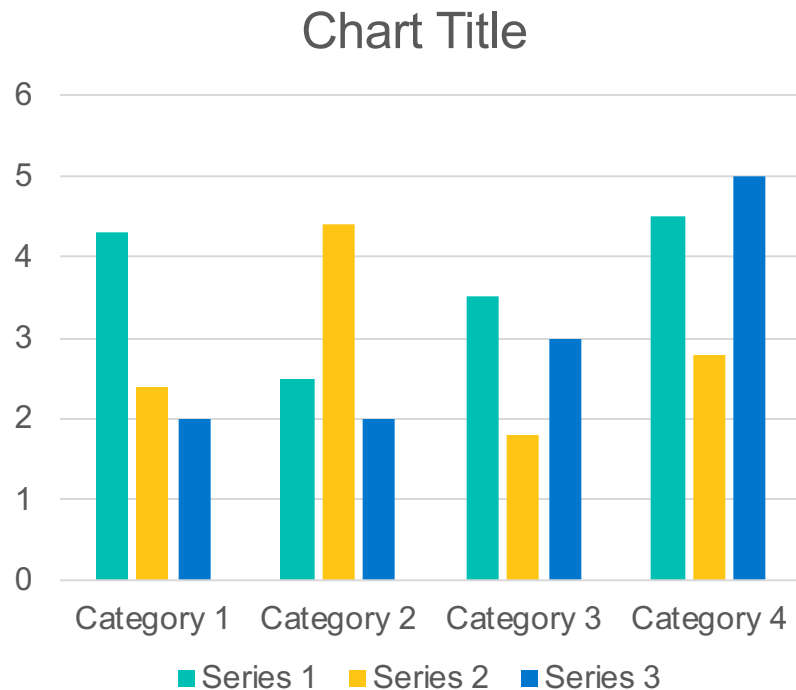
ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化

Dynamic PageMove Solution



ElasticSearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化

Dynamic PageMove Solution 性能数据



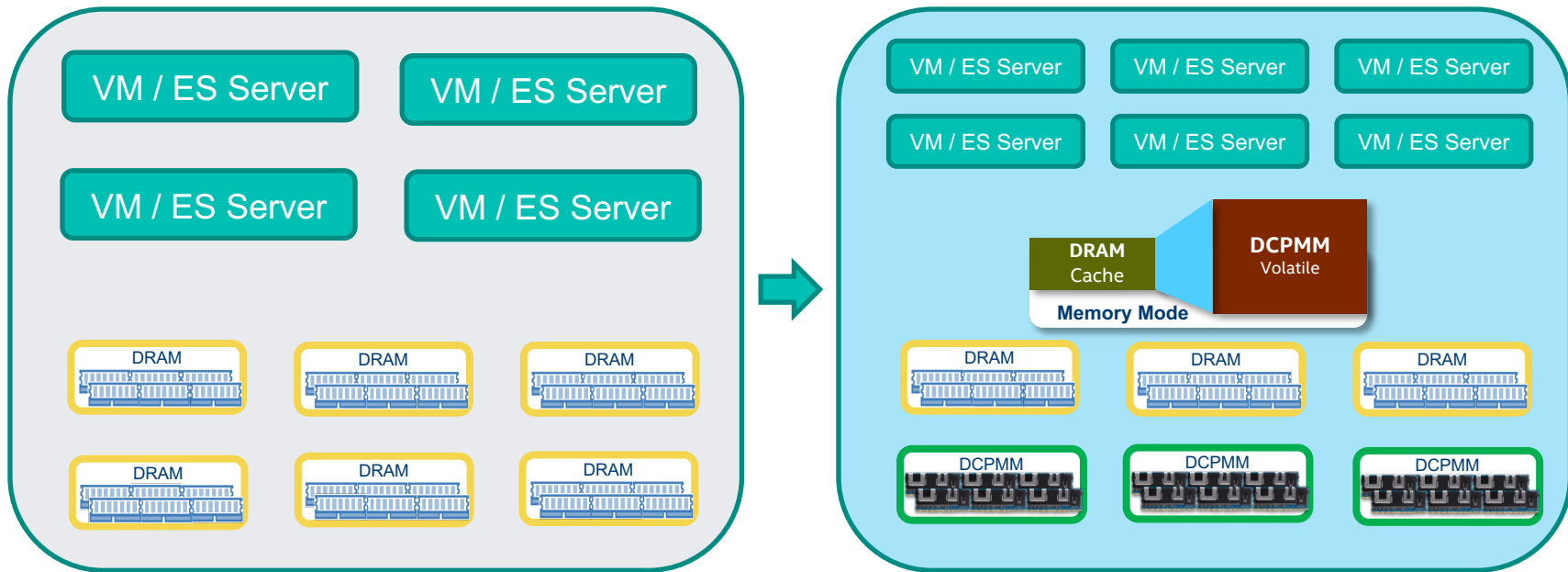
议程

英特尔® 傲腾™数据中心级持久内存Elasticsearch云的应用与实践

- 1 英特尔® 傲腾™数据中心级持久内存介绍
- 2 Elasticsearch 内存利用模式
- 3 Elasticsearch Java堆内内存利用英特尔傲腾数据中心级持久内存的优化
- 4 Elasticsearch Java堆外内存利用英特尔傲腾数据中心级持久内存的优化
- 5 在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化**

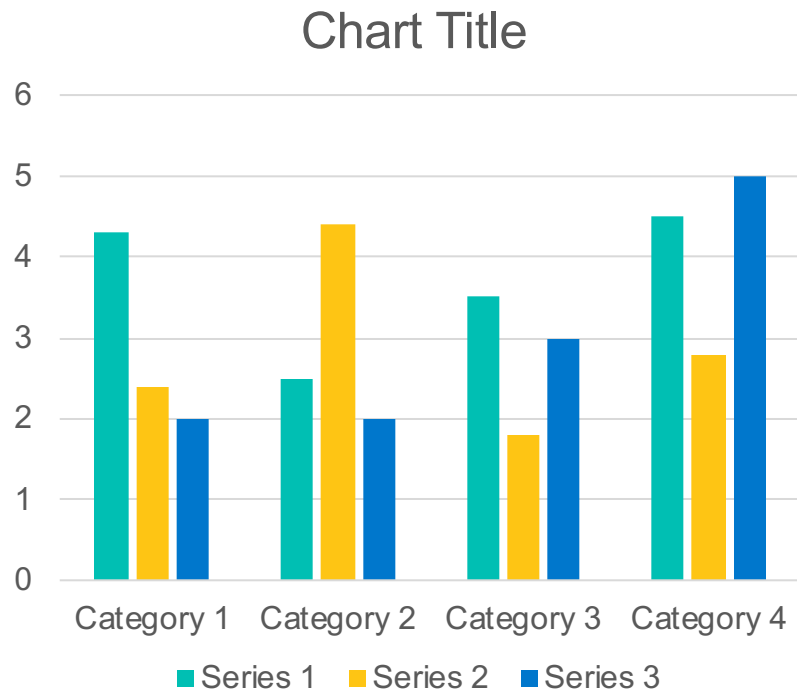
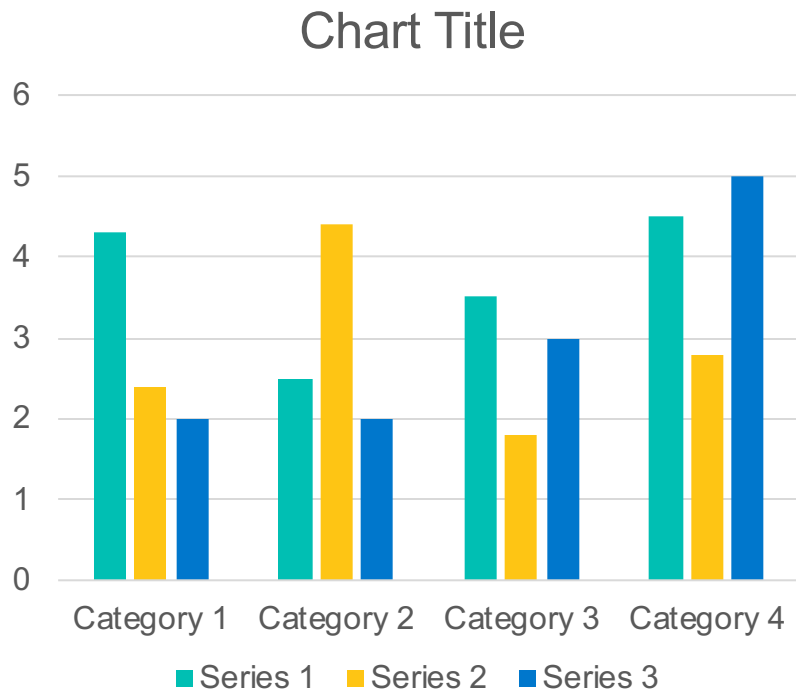
在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

- 使用Memory Mode增大主机可用内存，从而支持更多的虚拟机
- 软件系统不需要修改



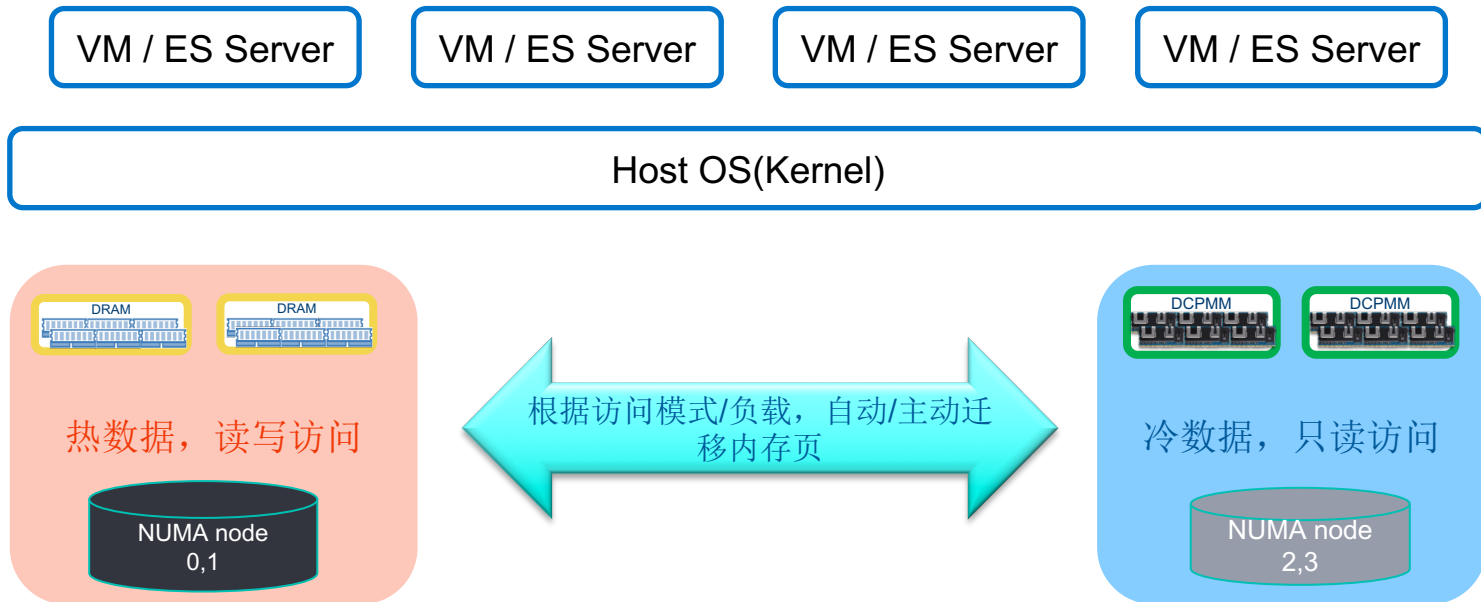
在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

Memory Mode 性能数据



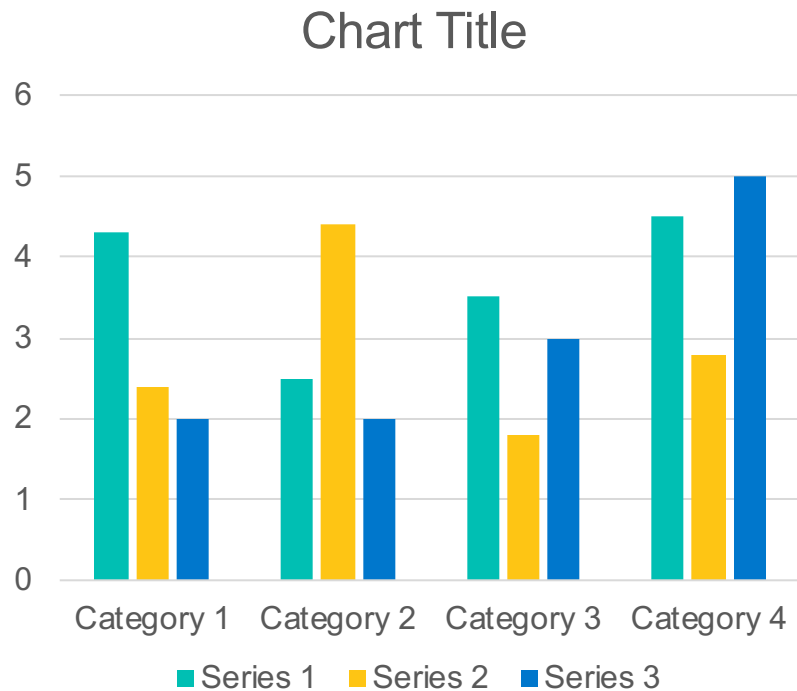
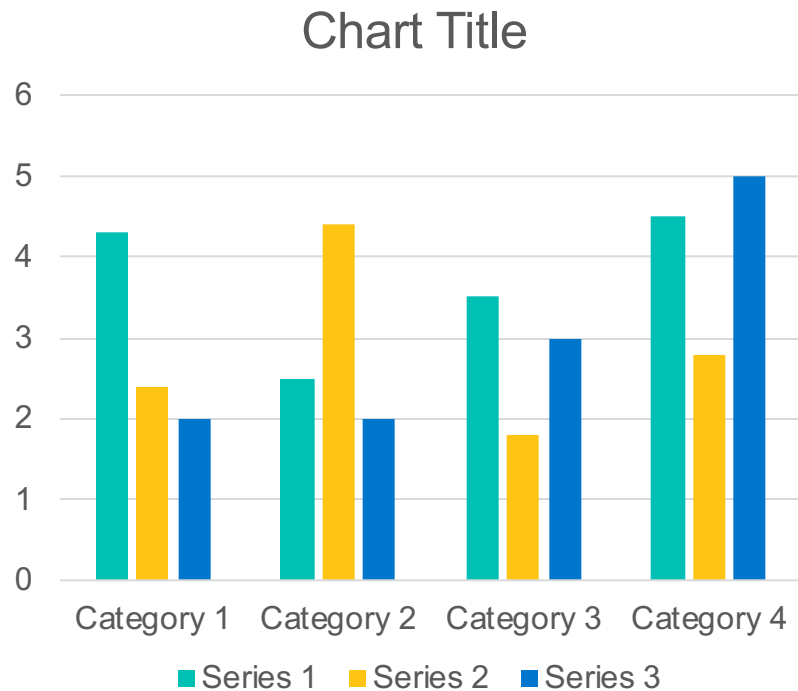
在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

- 使用App Direct Mode，新创建基于DCPMM的NUMA节点
- 内存数据根据访问模式在DDR节点和DCPMM节点间迁移
- 充分利用内存容量：总容量 = DDR + DCPMM
- 更高的灵活性：可以主动选择数据何时放在何处



在虚拟机上的ElasticSearch利用英特尔傲腾数据中心级持久内存的优化

App Direct Mode 性能数据



谢谢
Thanks