CASE STUDY

Intel® Optane™ DC Persistent Memory
Short Videos

Kuaishou's Recommendation System and Redis Services Have Storage Upgraded with Intel® Optane™ DC Persistent Memory to Complement DRAM

快手
记录世界 记录你

As a leading platform provider for short videos, Kuaishou is winning popularity among massive users for its real-time video content with great efficiency and precision. With the explosive increase of users and short videos, Kuaishou needs to introduce more advanced technologies constantly to tune and optimize its system architecture. And its storage system, as a core component for storage, distribution and recommendation in the short-video system, is also facing immense challenges in terms of optimization and performance enhancement.

In response to the challenges brought by the application scenarios with high throughput and large data requests in the short video app, Kuaishou has teamed-up with Intel for an in-depth technological collaboration: it is among the first to apply Intel® Optane™ DC persistent memory to its recommendation system and Redis service. In addition, with software fine-tuning and optimization, Kuaishou has successfully constructed a brand-new heterogeneous recommendation storage system and optimized Redis service to deliver greater storage capacity.

Kuaishou's tests and practices show that Intel Optane DC persistent memory used in the new heterogeneous recommendation storage system and upgraded Redis service not only have similar performance to DRAM but also enables better availability of the system with its high capacity and non-volatility. Furthermore, its advantages over DRAM in terms of cost and capacity also help Kuaishou reduce the total cost of ownership (TCO) of its recommendation system and Redis service.

Benefits of Kuaishou's solution:

- Kuaishou's heterogeneous recommendation storage system built on Intel Optane DC persistent memory not only fulfills core performance indicators such as request volume, network bandwidth and average processing latency, it also offers additional advantages in terms of capacity and cost compared to DRAM-based solutions
- Non-volatility of Intel Optane DC persistent memory enables better availability of Kuaishou's recommendation system, shortening its failure recovery time by up to a hundredfold[1]
- Kuaishou's heterogeneous storage system with Intel Optane DC persistent memory helps lower TCO of Kuaishou's recommendation system by 30% while meeting the performance requirements[2]
- With Intel Optane DC persistent memory, Kuaishou has more than doubled its memory capacity for a single Redis instance. TCO for its Redis service has also been reduced by 30%[3]

Reconstructing the storage system for the mega-scale recommendation system

The growing popularity of short videos on the internet is attracting more people to engage in producing and sharing short videos with appropriate apps. As a leading platform provider for short videos in China, Kuaishou has 200 million daily active users and tens of millions of short video uploads every day[4]. As such, when constructing the back-end system, it became one of Kuaishou's main foci to bridge its massive users and massive short video content, which would enable more users to load videos of their preferences on screen in real-time, and comment on, 'like' or 'dislike' the content at any time.

In response to users' demands for the real-time recommendation of video content, Kuaishou has been investing a large amount of resources in the construction and technological update of its content recommendation system since its inception. With the increasing number of users and short videos, it is key for Kuaishou to recommend suitable content to different users from its databases with tens of billions of short videos using a deep learning model with hundreds of billions of parameters, while supporting hundreds of thousands of concurrent calls per second at peak times. To do this, Kuaishou is embracing the latest technological trends and has constructed a recommendation system architecture with separated computing and storage based on heterogeneous devices.

As shown in Fig. 1, Kuaishou adopts an architecture with separated computing and storage in its recommendation system. It is composed of computing services (e.g. recommendation, prediction and recall) and storage services (e.g. user profiles, the parameter server and distributed indexing). The former is responsible for work such as recommendation strategy computing, model prediction and video retrieval, while the latter offers storage and real-time updating capabilities for hundreds of millions of user profiles, billions of short video features and hundreds of billions of ranking model parameters in the recommendation system.

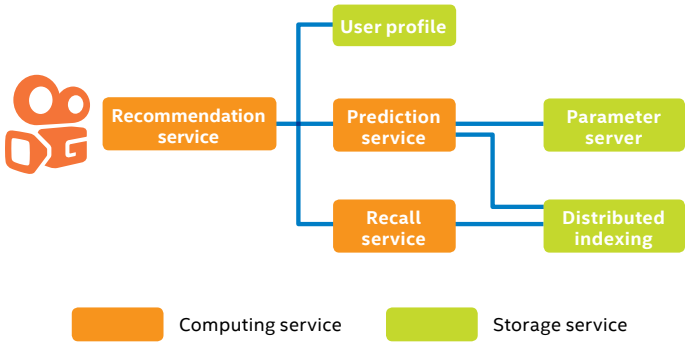

Fig. 1 Kuaishou's recommendation system architecture with separated computing and storage

A well-known typical application scenario for short videos is fragmented time. As users surf randomly through swipes on the Kuaishou app, the time for the recommendation system to process data is often within milliseconds. In addition to offering high-performance strategy computing in the computing module, it is undoubtedly more challenging to enable hundreds of millions of stored data to provide real-time support for the recommendation system.

As such, Kuaishou adopts diverse ways of implementing technology based on heterogeneous equipment according to different application scenarios. Take distributed indexing as an example: the power of indexing is crucial for high-speed data retrieval in large-scale distributed storage clusters. To enhance indexing performance under high concurrency, Kuaishou adopts memory-based key value (KV) databases to construct its distributed indexing system.

The performance of Kuaishou's Redis service, as another important cornerstone of the recommender system, has a significant influence on the given recommendations. Users' history of behaviors in the short video app is stored in the Redis database and used eventually to construct precise user profiles. The larger the memory capacity that can be used for Redis instances, the more information can be stored within the memory for high-speed access. As a result, user profiles will be more specific, enabling more precise recommendations to individual users.

Furthermore, the Redis service also strongly supports the social interactions (e.g. 'like', commenting and bullet screens) of Kuaishou short videos. The memory-based Redis database ensures the smooth operation of these social activities for excellent user experience.

However, with the rapid increase of data, Kuaishou's memory-based recommendation storage service and Redis service face increasing challenges. On one hand, the limited capacity of DRAM in a physical server makes it difficult to scale up memory for various service instances. On the other hand, expensive DRAM also significantly increases Kuaishou's TCO. The volatility of DRAM also results in more time for the system to recover from a failure.

To overcome these challenges and continue to provide users with a better content recommendation service, Kuaishou has teamed-up with Intel for in-depth technological collaboration, in addition to using heterogeneous mixed computing solutions to enhance the performance of its computing services. By introducing Intel Optane DC persistent memory, Kuaishou has optimized and transformed its recommendation storage system and Redis databases.

Complementary software and hardware for greater storage capacity

In traditional storage architecture, large-capacity persistent storage is mainly implemented using hard disk drives (HDDs) or solid state drives (SSDs). With increasingly diverse data application scenarios and more demanding requirements for storage performance, the hierarchy of storage requirements is becoming increasingly complex. The use of more DRAM will undoubtedly enable stronger performance, but it also brings higher costs. To resolve this, Kuaishou chose to build a brand-new heterogeneous storage structure for optimized performance, capacity and cost.

In Kuaishou's original design, high-performance DRAM is used for storage workloads requiring the highest performance but the least capacity, while SSDs and HDDs are used for persistent storage workloads requiring low performance requirements but high capacity requirements. However, Kuaishou still had to face another possible scenario: what if the storage system has high requirements for performance, capacity and persistency all at the same time?

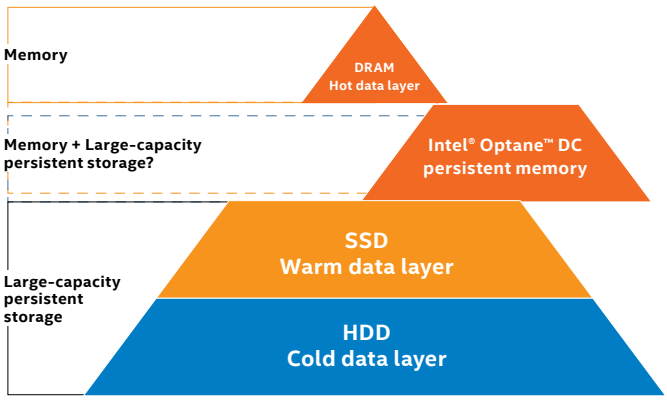

Fig. 2 Intel® Optane™ DC persistent memory is an ideal choice for both memory performance and large-capacity persistent storage

As shown in Fig. 2, the 3D XPoint™ storage medium-based Intel Optane DC persistent memory is an ideal choice for Kuaishou to fill the gap. Not only does this innovative product line of memory have similar read/write performance and access latency to DRAM, and higher durance than SSD, but it also enables near-DRAM performance in a highly concurrent recommendation system scenario. In addition, it enables Kuaishou to build a TB-level memory database with its large memory capacity. More importantly, it offers data persistency (or non-volatility, in the App Direct mode) that DRAM does not have, allowing greater availability of Kuaishou's heterogeneous recommendation storage system.

To maximize the performance of the heterogeneous storage system composed of DRAM, Intel Optane DC persistent memory, SSDs and HDDs, Kuaishou works alongside partners such as Intel to conduct feasibility analysis and architectural design research for different scenarios that its recommendation system may face. At the same time, it redesigns KV storage in distributed indexing and parameter servers based on the features of Intel Optane DC persistent memory.

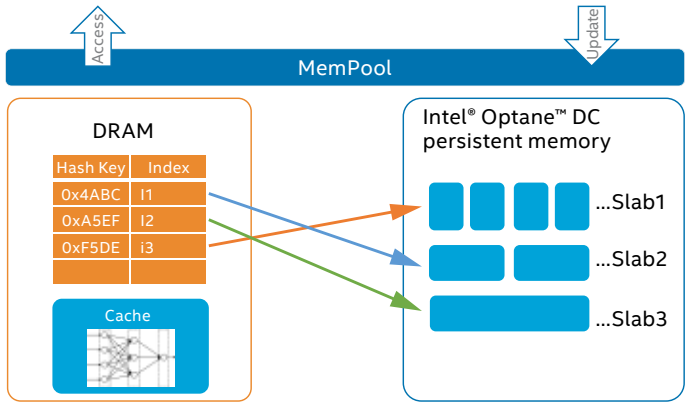

Fig. 3 Heterogeneous storage system built on Intel® Optane™ DC persistent memory

The new design is shown in Fig. 3, with the MemPool component added in the system architecture. As a cache pool, this component enables the system to determine whether to use DRAM or Intel Optane DC persistent memory according to different access types. For instance, when a parameter server is used for recommendation model prediction, the neural network can be allocated to DRAM by MemPool to enhance the prediction performance as its size is smaller than embedding tables in the model. And in the use case of distributed indexing, the system will allocate different slabs (a memory distribution mechanism) for it in Intel Optane DC persistent memory according to the size of the required indexed data to improve performance and efficiency of data access.

Apart from these major designs, Kuaishou has also implemented fine-tuning and optimization solutions based on the features of Intel Optane DC persistent memory. Firstly, for data access: binding Non-Uniform Memory Access Architecture (NUMA) nodes is used to avoid Intel Optane DC persistent memory's switching between different NUMA nodes when accessing data so that better read/write performance can be achieved. Additionally, the inclusion of Lock-Free and Zero Copy technologies also helps to prevent frequent access of critical sections to Intel Optane DC persistent memory and reduce memory bandwidth usage of data access, enhancing the storage system's overall performance. Meanwhile, non-volatility of Intel Optane DC persistent memory enables the newly designed indexing system to recover from a failure in minutes, which is a hundredfold increase compared to the recovery that took hours in the past[1].

As for the Redis service, the large-capacity Intel Optane DC persistent memory enables the large TB-level memory capacity of Kuaishou's single Redis server, expanding the memory capacity of a single Redis instance from 4GB to 8GB. By doubling the memory capacity of an instance, it lays a stronger hardware foundation for Kuaishou's further development in business operations.

Reducing TCO while fulfilling performance requirements

To verify the actual performance of Kuaishou's brand-new heterogeneous storage structure after adopting Intel Optane DC persistent memory and implementing a series of software optimizations, Kuaishou and Intel leveraged real-world online request data to conduct simulated pressure tests on relevant Intel Optane DC persistent memory-based systems including the indexing system used in the recommendation system.

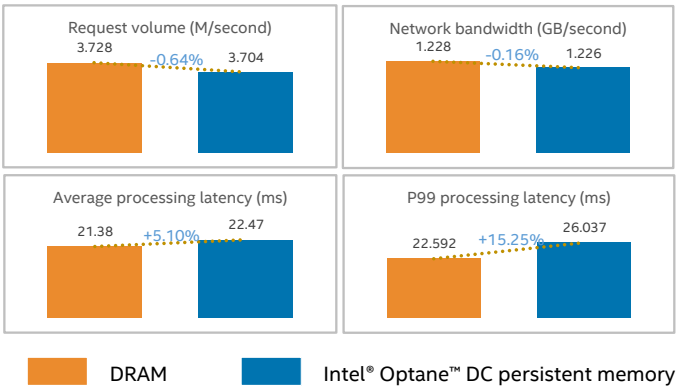

Fig. 4 Pressure test results of the indexing system based on Intel® Optane™ DC persistent memory

These tests were conducted around the recommendation system's four core performance indicators: request volume, network bandwidth, average processing latency, and P99 processing latency. The results are as shown in Fig. 4.

It is clear to see that Intel Optane DC persistent memory has similar performance to DRAM for the four core indicators, and the difference between these two is only 0.16% in terms of the network bandwidth indicator[5].

Intel Optane DC persistent memory featuring larger capacity, non-volatility and greater affordability compared to DRAM allows Kuaishou to control its costs effectively while delivering similar performance. Kuaishou's estimation shows that the introduction of Intel Optane DC persistent memory has reduced the TCO for its recommendation system and Redis service by 30%[2,3].

Conclusion

As one of the first Internet enterprises in China that introduced Intel Optane DC persistent memory to its recommendation system, Kuaishou, with its excellent technological innovation capabilities, has conducted meaningful exploration into the construction and application of a heterogeneous storage structure in its recommendation system as well as into the application of a large-capacity Redis service for short video services. Meanwhile, such exploration has also achieved great results.

Looking forward, Kuaishou is exploring the possibility of establishing a joint laboratory with Intel to drive its business innovation and the upgrading and evolution of its data centers. This application of Intel Optane DC persistent memory is the first project implemented while preparing for the establishment of the laboratory. In the future, Kuaishou will continue its cooperation with Intel to explore the application values of Intel Optane DC persistent memory in other business scenarios or services, promoting optimization and transformation of its various data processing and storage systems.

[1] Source of data: https://36kr.com/p/5232799
[2] Source of data: https://36kr.com/p/5232799
[3] Cost results are based on Kuaishou's internal measurement. For more details, please contract Kuaishou.
[4] Source of data: https://36kr.com/p/5232799
[5] Test results are based on Kuaishou's internal tests and evaluation. For more details, please contact Kuaishou.