HDFS Persistent Memory Cache Performance Test

feilong.he@intel.com
Summary

HDFS Persistent Memory Cache performance

- **DFSIO**
  - HDFS Persistent Memory Cache delivers **12.53X (random read), 22.06X (sequential read)** speedup compared to no cache (HDD storage).
  - HDFS Persistent Memory Cache delivers **4.95X (random read), 8.14X (sequential read)** speedup compared to DRAM cache.

- **Decision Support Workload - 1TB**
  - HDFS Persistent Memory Cache delivers **5.60X** speedup compared to no cache (HDD storage) for text format data.
  - HDFS Persistent Memory Cache delivers **2.67X** speedup compared to DRAM cache for Text format data.

- **Decision Support Workload - 2TB**
  - HDFS Persistent Memory Cache delivers **1.88X and 1.49X** speedup compared to no cache (HDD storage) respectively for Parquet and ORC format data.
  - HDFS Persistent Memory Cache delivers **1.3X and 1.07X** speedup compared to DRAM cache respectively for Parquet and ORC format data.
## Test Configuration (same cost for DRAM and PM)

<table>
<thead>
<tr>
<th></th>
<th>DRAM</th>
<th>HDD (no cache)</th>
<th>Persistent Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Memory</strong></td>
<td>DDR4 dual rank 768GB = 24 * 32GB</td>
<td>DDR4 dual rank 192GB=12 * 16GB</td>
<td>DDR4 dual rank 192GB=12 * 16GB Persistent Memory 8 * 128GB ES2</td>
</tr>
<tr>
<td><strong>Cluster</strong></td>
<td>2-node cluster</td>
<td>2-node cluster</td>
<td>2-node cluster</td>
</tr>
<tr>
<td></td>
<td>1 Namenode and 2 DataNode</td>
<td>1 Namenode and 2 DataNode</td>
<td>1 Namenode and 2 DataNode</td>
</tr>
</tbody>
</table>

Test Configuration (same cost for DRAM and PM)
Test case: DFSIO Read

Workload
- DFSIO (Hadoop’s Distributed I/O Benchmark)
  Random & Sequential Read

Total Data Set
- 128 * 8GB = 1024GB

Persistent Memory cache capacity
- 920G (almost fully cached)

DRAM cache capacity
- 570GB (partially cached)

Metrics
- Throughput (MB/s)

Baseline
- 6 * HDD (no cache)
DFSIO Performance Overview

- Persistent Memory (PM) cache delivers **4.95X (random read), 8.14X (sequential read)** speedup compared to DRAM cache.
- Persistent Memory cache delivers **12.53X (random read), 22.06X (sequential read)** speedup compared to no cache.
- Persistent Memory brings more performance improvement for Sequential Read than Random Read.
- When dataset cannot fit memory cache pool, DRAM cache’s performance is much worse than Persistent Memory’s.
Test case: Decision Support Workload 1TB

Workload
1TB dataset for 54 selected queries in different format: Text, Parquet, ORC

Total Data Set:
- 1TB Parquet (real size: 422GB)
- 1TB ORC (real size: 359GB)
- 1TB text (real size: 916GB)

Persistent Memory cache capacity
920GB

DRAM cache capacity
570GB

Metrics
Query time (sec.)

Baseline
6 * HDD (no cache)
Decision Support Workload 1TB

Caching tables into HDFS Cache
- Added sequentially by table file name
- HDFS tries to cache as much as possible

Table details

- **25 tables referred by 54 queries**
  - Text format: DRAM cannot fully cache store_sales (390GB), web_return (10GB), web_sales (147GB)
  - Parquet & ORC format: both Persistent Memory and DRAM can cache all tables.

- **10 tables referred by 7 IO-Intensive queries**
  - Text format: DRAM cannot fully cache store_sales (390GB)
  - Parquet & ORC format: both Persistent Memory and DRAM can cache all tables.

### Num. of Tables Cached for 54 Queries

<table>
<thead>
<tr>
<th>Data Format</th>
<th>Persistent Memory</th>
<th>DRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parquet</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>ORC</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Text</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

### Num. of Tables Cached for 7 IO-intensive Queries

<table>
<thead>
<tr>
<th>Data Format</th>
<th>Persistent Memory</th>
<th>DRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parquet</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>ORC</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Text</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
• Persistent Memory (PM) Cache provides **2.67x** and **5.60X** speedup over DRAM and HDD for text format data in Decision Support (DS) Workload.
  • Persistent Memory can cache all the tables, DRAM cannot.

• Persistent Memory cache performance is lower (10%-14%) than DRAM in parquet & orc format.
  • Data can be fully cached into Persistent Memory and DRAM due to data compression.
Performance Overview for 7 IO Intensive Queries

In general, Persistent Memory brings higher performance improvement for I/O intensive queries.
Test case: Decision Support Workload 2TB

Workload
2TB dataset for 54 selected queries

Total Data Set
- 2TB Parquet (real size: 816GB)
- 2TB ORC (real size: 706GB)

Persistent Memory cache capacity
920GB (fully cached)

DRAM cache capacity
570GB (partially cached)

Metrics
Query time

Baseline
6 * HDD (no cache)
**Caching tables into HDFS Cache**
- Added sequentially by file name
- HDFS tries to cache as much as possible

**Table details**
- **25 tables referred by 54 queries**
  - DRAM cannot fully cache all tables
  - Persistent Memory can cache all tables.
- **10 tables referred by 7 IO-intensive queries**
  - DRAM cannot fully cache table `store_sales`. ~12GB (~4%) not cached in ORC, and ~107GB (~30%) not cached in Parquet.
  - Persistent Memory can cache all tables.

---

<table>
<thead>
<tr>
<th>Data Format</th>
<th>Persistent Memory</th>
<th>DRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parquet</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>ORC</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Data Format</th>
<th>Persistent Memory</th>
<th>DRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parquet</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>ORC</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
Performance Overview for 54 Selected Queries (2TB)

- Persistent Memory (PM) Cache provides **1.88X (Parquet)** and **1.49X (ORC)** speedup over no cache respectively in Decision Support (DS) Workload.
- Persistent Memory Cache provides **1.3X (Parquet)** and **1.07X (ORC)** speedup over DRAM.
  - Persistent Memory can cache all the tables and DRAM cannot.
  - DRAM cached more tables on ORC than on Parquet.

Persistent Memory provides higher performance gain for Decision Support Workload in larger dataset
Notices and Disclaimers

• No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

• Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

• This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

• The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

• Intel, the Intel logo, Xeon, Optane, Optane DC Persistent Memory are trademarks of Intel Corporation in the U.S. and/or other countries.

• *Other names and brands may be claimed as the property of others

• © Intel Corporation.
Legal Information: Benchmark and Performance Disclaimers

• Performance results are based on testing as of Feb. 2019 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.

• Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information, see Performance Benchmark Test Disclosure.

• Configurations: see performance benchmark test configurations.