Software Defined Memory
Group Update

Manoj Wadekar, Meta
SDM Team Charter

- **Identify** key applications driving adoption of Hierarchical/Hybrid memory solutions
- Establish architecture and **nomenclature** for such Systems
- Offer **benchmarks** that enable validation of novel ideas for HW/SW solutions for such systems
SW Defined Memory (SDM) Group

- SDM opportunity is around data
  - The growth in data and its processing has ramped quickly in OCP data centers
  - In-memory data processing allows processing in parallel and real-time for insights
- Challenge is better usage of memory for data workloads
  - Memory used for data processing is being asked to do more (bigger, faster & more resilient)
  - Physical memory options are proliferating allowing use case innovation
- SDM abstractions are expected to release new value for memory use
  - Expected usage: location transparency, protocol abstraction, automated tiering, unified namespace, thin provisioning and fine grained QoS

New memory markets can be enabled by merging software and novel memory technologies
SDM investigation roadmap

Team & charter
- Definition document
- Definition slides
- SDM Survey
- OCP use case inputs
- FTI – Industry presentations

Publish Foundational doc
- Work (1-3) use cases
- Expand circle of expertise
- Establish thought leadership

Demonstrate (1-3) use cases
- Facilitate demonstrations
- Evangelize work
- Get industry involved
- Work additional use cases (according to feedback)

OCP Project for SDM
- Fold project back into OCP
- Establish incubator project

Investigation approach:
- Identify opportunities in industry to improve memory use
- Understand workload problems and select use cases
- Develop solutions use case architectures
- Offer Benchmarks for work-load specific use cases
Survey

Identify key applications driving adoption of Hierarchical/Hybrid memory solutions
Modern applications demand memory expansion, driving needs vary.
Motivation for Software Defined Memory
Obtained via Industry Survey

- OS Managed: 43%
- Application Managed: 28%
- Hardware Managed: 24%
- No Preference: 5%

Strong interest in Software Managed Memory solution

Is pooling and sharing of memory capacity across multiple servers as a key aspect for your future infrastructure growth?

- Extremely Important
- Somewhat Important
- Not...

Interconnect Technologies of Interest in the Next 2-3 Years

- Memory on a Multi-Host Fabric such as Gen-Z
- Memory on CXL Switch-Based Fabric
- Memory on CXL Server Interconnect

Memory Expansion needs: Software managed, server-local and pooled
SDM Nomenclature

Establish architecture and nomenclature for SDM Systems
Software Defined Memory (SDM) is an emerging architecture paradigm that provides software abstraction between applications and underlying memory resources with dynamic memory provisioning to achieve the desired application SLA.
SDM Hierarchy – Logical View

- **1\textsuperscript{st} Level Memory**: Preferred for OS and/or application memory allocation
  - Socket local DRAM
  - UPI connected DRAM
  - HBM

- **2\textsuperscript{nd} Level Memory**: Option for application memory allocation based on SLA
  - DDR-\textit{x} connected Storage Class Memory
  - Memory expansion with CXL connected DRAM or SCM
  - Compressed memory pages on block storage
  - CXL fabrics based pooled memory

Differentiated attributes:
- Bandwidth and Latency
- Cost and Power
- Volatility versus Persistence
- Computational offload characteristics

CPU/GPU

1\textsuperscript{st} Level Memory

2\textsuperscript{nd} Level Memory

Block Storage

Pagefile
2nd level memory use cases:

- Compressed page store in zSwap based memory for efficiency improvements
- Bandwidth memory expansion for Memcache class applications
- Capacity memory expansion for database IO cache & user mode memory allocations
- Application aware Persistent Memory integrations
SDM – Software View

Kernel managed placement

1st Level Memory

2nd Level Memory

User Space, Kernel Space

malloc

Application

App managed placement

1st Level Memory

2nd Level Memory

malloc
Offer benchmarks that enable validation of novel ideas for HW/SW solutions for such systems
## SDM Applications, benchmarks

<table>
<thead>
<tr>
<th>Category</th>
<th>Workload</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Micro benchmarks** | Intel® Memory Latency Checker (MLC)  
https://www.intel.com/content/www/us/en/developer/articles/tool/intel-memory-latency-checker.html | Measure memory latencies and b/w, and how they change with increasing load on the system. |
http://www.cs.virginia.edu/stream/ref.html | simple, synthetic benchmark designed to measure sustainable memory bandwidth (in MB/s) and a corresponding computation rate for four simple vector kernels: Copy, Scale, Add and Triad |
| **fio (Flexible I/O Tester)** | https://github.com/axboe/fio.git | Block, pmem benchmark tool |
| **Caching** | CacheBench  
https://cachelib.org/docs/Cache_Library_User_Guides/Cachebench_Overview  
https://github.com/facebook/CacheLib.git | benchmark and stress testing tool to evaluate cache performance with real cache workloads |
| **memtier** | https://github.com/RedisLabs/memtier_benchmark | Redis and Memcache traffic generation and benchmarking tool |
## SDM Applications, benchmarks

<table>
<thead>
<tr>
<th>Category</th>
<th>Workload</th>
<th>Descrit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Databases</strong></td>
<td>db_bench</td>
<td>benchmark RocksDB's performance</td>
</tr>
<tr>
<td></td>
<td><a href="https://github.com/facebook/rocksdb.git">https://github.com/facebook/rocksdb.git</a></td>
<td></td>
</tr>
<tr>
<td>SysBench</td>
<td><a href="https://github.com/akopytov/sysbench">https://github.com/akopytov/sysbench</a></td>
<td>Database benchmarking (e.g. MySQL)</td>
</tr>
<tr>
<td>HammerDB</td>
<td><a href="https://hammerdb.com/">https://hammerdb.com/</a></td>
<td>benchmarking and load testing software for databases (Oracle Database, SQL Server, IBM Db2, MySQL, MariaDB and PostgreSQL).</td>
</tr>
<tr>
<td><strong>AI</strong></td>
<td>MLPerf</td>
<td>performance benchmarks that cover a range of leading AI workloads widely in use.</td>
</tr>
<tr>
<td></td>
<td><a href="https://github.com/mlcommons/inference">https://github.com/mlcommons/inference</a></td>
<td></td>
</tr>
<tr>
<td><strong>Bigdata</strong></td>
<td>HiBench</td>
<td>big data benchmark suite</td>
</tr>
<tr>
<td></td>
<td><a href="https://github.com/Intel-bigdata/HiBench">https://github.com/Intel-bigdata/HiBench</a></td>
<td></td>
</tr>
</tbody>
</table>
Benchmarks - gaps

- Transparent Memory use cases can take advantage of existing benchmarks
- We need benchmarks that are focused on tiered memory
  - E.g. Kernel tiering – how to track hot/cold pages - perf impact
- Application changes may be required for taking advantage of App-managed-memory
- Need industry’s help
CacheBench

- **CacheLib** – pluggable **in-process caching engine** to build and scale high-performance services
  - C++ Library
  - Thread-safe API
  - Manages DRAM and Block Caching transparently

- Meta open-source project: [https://github.com/facebook/CacheLib](https://github.com/facebook/CacheLib)
  - See [www.cachelib.org](http://www.cachelib.org) for documentation and more information.

- **CacheBench** - benchmarking tool for evaluating caching performance
Next Steps

Team & charter
- Definition document
- Definition slides
- SDM Survey
- OCP use case inputs
- FTI – Industry presentations

Publish Foundational doc
- Work (1-3) use cases
- Expand circle of expertise
- Establish thought leadership

During '21
- Feb '21
- FTI '21

Spring '22
- FTI '22

OCP Project for SDM
- Fold project back into OCP
- Establish incubator project

During '21
- Spring '22
- Fall '22
- Beyond

Investigation approach:
- Identify opportunities in industry to improve memory use
- Understand workload problems and select use cases
- Develop solutions use case architectures
- Offer Benchmarks for work-load specific use cases

OCP Future Technologies Symposium
Call To Action

● Help define finalizing Storage Defined Memory architecture and use cases
● Help define benchmarks for the key use cases using SDM
● Join the team: https://www.opencompute.org/projects/software-defined-memory-workstream
Thank You
SDM scenario in GPU training

- **Usage Examples:**
  - GPUs accessing CPU attached memory through DMA
  - GPU accessing NVMe block storage for training parameters
  - Other current and future interconnects: CXL, NVLink etc.
SDM Hierarchy – SW (NUMA) View

- Kernel can enumerate devices
  - 1LM device showing up on Node 0
  - All other devices on separate Nodes
- SCM can be accessed through DAX in system-ram mode
  - As remote NUMA
- Multiple Namespaces to tabulate devices