SCM – перспективы одноуровневой памяти

Преимущества SCM (Storage Class Memory), построенной на базе Persistent Memory (PMEM или nonvolatile memory – NVDIMM – энергонезависимая/постоянная память), которая уже в ближайшие годы (начиная с конца 2018 г.) кардинально изменит как архитектуру вычислительных систем, так и принципы разработки ПО.

Введение

Появление памяти класса SCM (Storage Class Memory) на базе PMEM расценивается многими аналитиками как одно из революционных событий десятилетия. SCM позволит не только удовлетворить требования все возрастающей доли современных нагрузок, связанных с:

- *аналитическими приложениями реального времени (включая большие данные, in-memory databases, NoSQL, BI, сложные запросы);*
- *виртуализацией / облачными вычислениями / IaaS (нововведения в Windows Server 2012 и 2016: расширении виртуализации с MS Hyper-V до 1024 VMs на хост и 64 vCPUs на VM с до 1TB RAM на VM и 64TB VHDX);*
- *высокопроизводительными HPC/GPU нагрузками;*
- *поддержкой контейнеров;*
- *открытыми программируемыми сетями (SDN/NFV)*

и др., но и качественно изменить подход к обработке данных, программированию, обеспечению целостности/сохранности/безопасности/комплайнсу данных.

SCM положительно скажется как с точки зрения появления новых бизнесов/сервисов, так и их ценовой доступности для всех уровней предпринимательства – от крупного до малого (рис. 1).

В качестве примеров преимуществ, которые можно получить от SCM сразу, можно отметить следующие:

- *повышение производительности для интенсивных вычислительных нагрузок (SCM позиционируется с задержкой 10–20 микросекунд вместо диапазона 100–200 микросекунд у флэш-накопителей.);*
- *установка большего объема SCM-памяти на серверах гораздо более экономична, чем увеличение количества дисков и массивов в SAN для повышения производительности базы данных;*
- *увеличение эффективности потребляемой мощности (в сравнении с HDD/SSD, особенно в случае LR-DIMM);*
- *улучшение планирования многоядерных процессоров;*
- *снижение свопинга виртуализированного приложения во время виртуализации, что увеличивает среднее число экземпляров виртуальных машин, которые могут выполняться на каждом хосте (иногда удваивая объем рабочей нагрузки).*

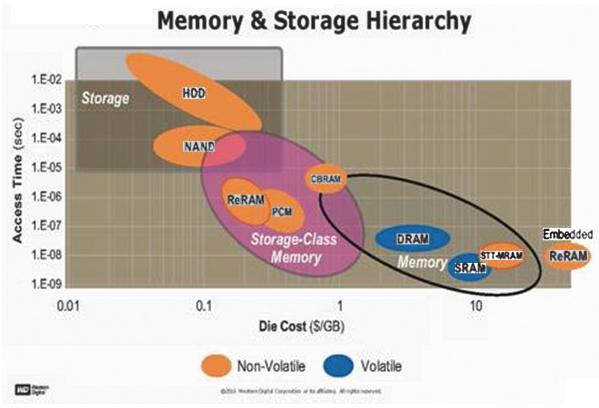

Рис. 1. SCM повысит и производительность приложений и их экономическую эффективность.

Тенденции в технологиях

Технологии и архитектуры традиционных хранилищ данных и оперативной памяти все в большей степени переплетаются. Тому свидетельство:

- *хранилище, построенное на полупроводниковых медиатехнологиях, таких как 3D NAND Flash, может достичь очень низких латентностей (<0,5 мс) и высоких IOP (измеряемое миллионами);*
- *слот DIMM становится все более популярным в качестве хранилища данных с непосредственным доступом процессора, предлагая доступ к данным с ультранизкой задержкой для расширенных вариантов использования, таких как энергонезависимая память (NVDIMM), либо для увеличения объема памяти сервера, либо для предотвращения HW & SW накладных расходов на канал передачи данных PCIe IO;*
- *storage-centric атрибут PMEM теперь является новой способностью ОП.*

Для полупроводниковых сред наиболее релевантной метрикой плотности является количество бит, которое может быть сохранено на отдельном чипе, а также количество чипов, которые могут вставляться на 300-миллиметровую кремниевую пластину. Сегодня самая передовая технология DRAM позволяет хранить 16 Гбит/с на чипе, а современная технология Flash – 512 ГБ. Эта фундаментальная структура затрат объясняет большую часть соотношения затрат 10-20x между DRAM и Flash (рис. 2,3).

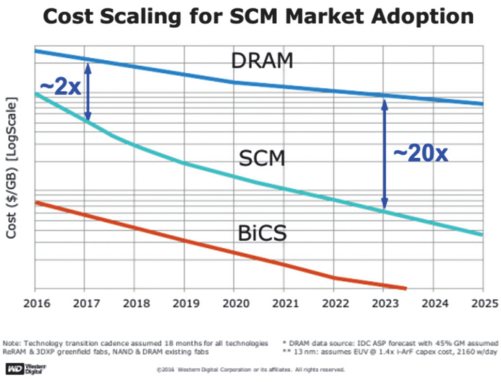

Рис. 2. Преимущества SCM над DRAM со временем будут возрастать.

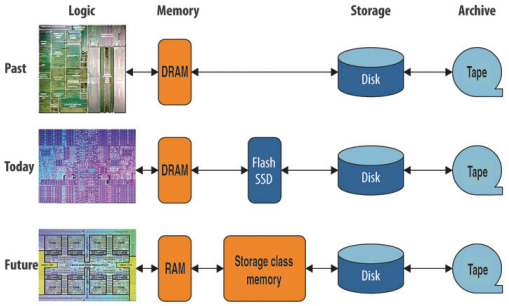

Рис. 3. Изменение во времени иерархии хранения данных.

Влияние SCM на отрасль

Влияние этих технологий будет огромным. Самым значительным воздействием на отрасль, вероятно, будет являться само свойство энергонезависимости памяти. Стойкие носители с временем доступа в десятки наносекунд позволят создавать целые новые классы систем и приложений. Теперь у разработчиков будет огромная, плоская, одноуровневая постоянная память как их устойчивая, последовательная основа данных. Разработчик приложения может свободно выбирать: как, когда и где хранить копию, размещать или не размещать ее где-то на запоминающем устройстве для безопасного хранения. С PMEM исчезают ограничения и ограничения волатильной DRAM – технология памяти больше не требует отдельной постоянной внешней копии. Копия может быть сделана для совместного ис-

пользования с другим компьютером в сети. Возможно, образ памяти будет проверен в файловой системе, которая будет бэкапироваться, архивироваться или обрабатываться любым из множества процессов на требования соответствия данных стандартам/рекомендациям. Теперь это явные, проактивные, независимые решения в области управления информационной политикой, а не побочные эффекты, возникающие в результате использования технологии энергозависимой памяти.

Разработчики приложений могут использовать преимущества постоянной памяти как фундаментальную особенность для написания приложений. Сегодняшние БД с in-memory-обработкой являются в основном эволюционными приложениями из мира чтения / записи данных. Хотя они были доработаны, чтобы использовать преимущества большой памяти, они обычно имеют резервную копию, которая делает необходимым хранение согласованных копий данных в составе приложения. Мы можем ожидать, что для мира PMEM будут созданы новые механизмы консистентности и восстановления.

История SCM

В рамках темы флэш-инвестиций появился набор новых возможностей памяти. Эти возможности «новой памяти» в основном подразделяются на две категории:

1. Fast Storage (быстрое хранилище). Это технологии, которые работают от 1-10 до 10-100 раз быстрее, чем флэш, для storage-приложений с высокой частотой обращений / низкой задержкой. Промышленность называет эти технологии «Storage Class Memory» (SCM). SCM часто используется как высокоскоростное устройство READ / WRITE для использования в качестве хранилища компьютеров и будет использоваться для создания твердотельных дисков (SSD). Технически, NAND Flash – это тип SCM.

2. Large & Non-Volatile Memory (большая и энергонезависимая память). Технологии, которые работают в области 50ns, попадают в эту категорию, как правило, в слот DIMM, но более высокоемкие и/или более рентабельные, чем DRAM. Эти технологии с «большой памятью» хорошо подходят для растущих потребностей нагрузок с высокой бизнес-ценностью и in-memory масштабируемостью. Кроме того, эти носители являются энергонезависимыми – они сохраняют свое состояние при отключении питания или перезагрузке, или неожиданном удалении, свойства сохранения данных, которые ранее были приписаны только устройствам хранения. Промышленным термином для этих технологий является «Постоянная память» (PMEM). PMEM чаще всего используется как устройство LOAD/STORE, используемое непосредственно процессорами в качестве основной памяти без вмешательства SW.

Архитектура SCM

Форм-фактор и интерфейс

ИТ-индустрия определила ряд стандартных электромеханических интерфейсов и протоколов данных, которые в целом обеспечивают независимые инновации и поддерживают взаимозаменяемость компонент между поставщиками компонент и системными решениями. В случае PMEM сегодня есть два основных интерфейса:

1. Форм-фактор DIMM и интерфейс DDRx: это основная модель памяти DRAM в экосистеме процессора x86. Архитектура процессоров и материнские платы определяют фиксированное количество каналов памяти DDR и слотов DIMM, предназначенных для соответствия ограничениям производительности, мощности и охлаждения. Некоторые решения SCM (Storage Class Memory), такие как 3DXP (Intel 3D Xpoint) от Intel, могут быть адаптированы для работы в рамках этих ограничений. В этом случае Intel создала проприетарное расширение для DDR для обеспечения более медленного времени доступа к 3DXP и DRAM. С точки зрения отрасли, идеальный PMEM будет работать в стандартной спецификации DDR без проприетарных или пользовательских расширений.

2. Интерфейс PCIe: это основной интерфейс ввода-вывода для процессора x86 и сегодня NVMe стал самым эффективным протоколом для поддержки блочного хранения на базе PCIe. Интересны 4 форм-фактора PCIe:

- *плата расширения (AIC, Add-In-Card)*, предназначенная для постоянной установки в физический слот PCIe;
- *U.2-устройство*, предназначенное для физически обратной совместимости с форм-фактором 2,5 "HDD;
- *Enterprise Datacenter Solid State Form Factor (EDSFF)* – новый форм-фактор, разработанный с нуля, для размещения и использования уникальных возможностей твердотельных носителей без "устаревшего багажа" для совместимости с форм-фактором жесткого диска. Dell EMC внедрила этот дизайн под названием «SoloFlex», при поддержке других производителей OEM-производителей и производителей SSD.

Кроме того, в рамках развиваемой датацентричной архитектуры сообществом производителей развивается протокол/стандарт GenZ, который становится новым типом фабрики, масштабируемой в рамках шкафа и который может быть заменен на DIMM/DDR и/или PCIe/NVMe *(см.: http://genzconsortium.org/ – для всестороннего изучения этой новаторской отраслевой технологии)*.

Семантика доступа

Информация, хранящаяся как "0" и "1", получает доступ к рабочим нагрузкам программного обеспечения одним из двух основных способов:

1. Read/Write (чтение/запись). Приложение, поддерживаемое библиотекой из фрэймворка или из языка программирования, обращается к своим данным, используя READ или WRITE системный вызов, семантика которого определяется программным интерфейсом POSIX. Этот стандарт интерфейса является основой мобильности приложений в течение почти 30 лет. Как правило, эти системные вызовы посылаются в ядро операционной системы, проходят через блок-уровень подсистемы хранения и приводят к операциям ввода-вывода, которые сохраняют или извлекают связанные данные с физического устройства хранения. Поскольку эти операции часто могут занимать несколько сотен микросекунд, операционная система отмечает, что это приложение ожидает завершения ввода-вывода, а затем переходит к поиску другой работы. Когда операция запоминающего устройства, наконец, завершается, операционная система прерывается достаточно долго, чтобы отметить, что ожидающая операция завершена и что приложение может быть возобновлено.

2. Load/Store (загрузка/сохранение). По сути, приложения состоят из небольших инструкций, которые кодируют логику программиста в последовательность двоичных или арифметических операций с данными. Современный процессор x86 выполняет эти операции путем быстрого доступа к данным непосредственно из своей основной памяти. Операция LOAD принимает данные из памяти и выводит их в логический движок процессора, чтобы выполнить некоторый шаг вычисления. Аналогично, операция STORE берет данные из процессора и помещает их в память. Процессоры, как правило, имеют небольшие, очень быстрые внутренние запоминающие устройства, известные как кэши, которые достаточно быстры, чтобы синхронно работать с процессором, – часто всего несколько наносекунд. Когда требуемый доступ не может быть обработан кэшем, процессор ждет, пока подсистема памяти завершит операцию, обычно в течение 50-100 наносекунд. Процессор ждет или «останавливается», пока операция памяти не будет выполнена.

SW-экосистема

1. Операционные системы. Windows и Linux быстро развиваются для создания или улучшения поддержки как операций с очень малой задержкой SCM IO (~ 1-10us READ / WRITE), так и высокоскоростных LOAD / STORE до PMEM. В результате IO SW стеки, файловые системы и менеджеры памяти ядра расширяются для поддержки нового класса приложений, желающих использовать эти новые возможности.

2. Приложения. Фрэймворки управления корпоративными данными оптимизированы для поддержки как очень быстрого ввода-вывода, так и очень большой процессорной памяти. SAP/HANA, Oracle/TimesTen and Microsoft SQL-Server In-Memory OLTP являются действующими примерами этой оптимизации. Кроме того, аналогично адаптируется множество разнообразных фрэймворков near / no-SQL и open-source. Техническая рабочая группа по программированию в области энергонезависимой памяти SNIA провела исключительную работу по подготовке как независимых разработчиков, так и IHV для этого нового мира. В частности, Энди Рудофф (Intel), возможно, является самым заметным и последовательным евангелистом в этой области.

3. Языки программирования. Большинство современных языков программирования и программистов понимают разницу между быстрым хранением (DAS/внеш-

ние СХД) и памятью (оперативной). Тем не менее, большинство из них также «знают», что хранилище (как правило) постоянно, а память (всегда) изменчива. Текущие языки программирования будут развиваться, и из академических кругов будут возникать новые модели программирования. Это не произойдет в одночасье, подобно длительному и изнурительному внедрению параллельных и многопоточных моделей программирования в ответ на экономические преимущества многоядерных процессоров.

4. Шифрование. Значительные отраслевые усилия пошли на поддержку шифрования Data-At-Rest. Накопители хранилищ хранилищ и твердотельные накопители были расширены для поддержки промышленных требований комплайнса к данным таких, как FIPS. С появлением PMEM эти модели должны быть расширены и на память. Предлагается несколько моделей на основе того, выполняется ли шифрование процессором, межсоединением памяти или внутри самого модуля памяти.

Области применений SCM, на которые ориентирована Dell EMC

Память версии 1.0, как известно, сложна и имеет существенные технологические риски. Среди рынков и компаний, готовых внедрять постоянную память, следующие:

- компании с гипермасштабируемой инфраструктурой, такие как Google, стремятся ускорить внедрение большой памяти серверов;
- раннее внедрение технологий PMEM Dell нацелено на замену большой памяти серверов и рабочих станций;
- платформы хранения, поскольку применение PMEM в них обосновано за счет того, что стеки программного обеспечения улучшаются, что дает возможность использовать PMEM для дифференцированных затрат, производительности и возможностей. PMEM, скорее всего, будет использоваться в разработке новых твердотельных накопителей, чтобы повысить производительность и избежать сложности современных встроенных DRAM и решений SuperCap;
- в долгосрочной перспективе использовать PMEM будут такие традиционные ISV, как SAP, ORCL и MSFT, чтобы использовать ценность сохранения данных, поскольку экосистемы PMEM SW созревают, и ожидается, что стартапы будут финансироваться в Office CTO, создавая новые неопасные, постоянные накопительные пакеты SW, где простота и доступность данных/вычислений являются первостепенными;
- наконец, архитектуры Edge / IoT HW, требующие локального захвата, хранения и анализа, извлекают выгоду из единой простой модели постоянной памяти с чрезвычайно низким энергопотреблением и мгновенным ответом.

Рынок SCM

Каждый из игроков (Intel / Micron, Toshiba / SanDisk, SK Hynix, Samsung) вкладывает значительные средства в «Fast Storage»:

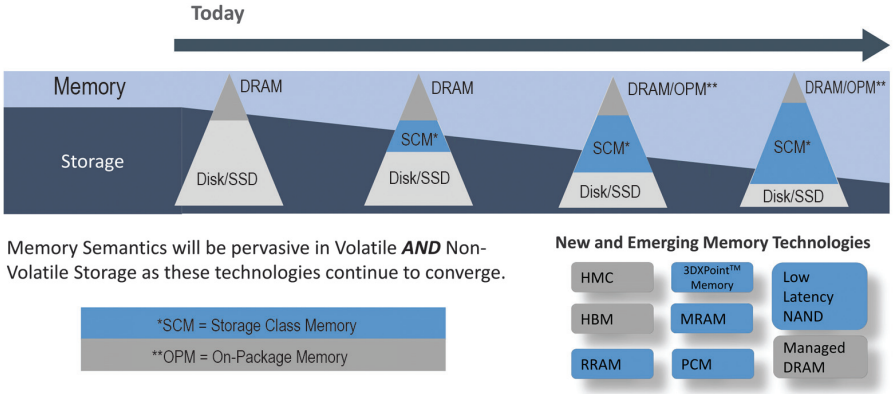

Рис. 4. Со временем доля SCM будет возрастать и замещать традиционные уровни хранения на HDD/SSD.

- 3D CrossPoint (3DXP) – Intel / Micron. Это технология резистивного ОЗУ (RAM) с фазовым переключением. Она быстрее, чем флэш, имеет бОльшую емкость, но медленнее, чем DRAM. В настоящее время Intel позиционирует эту технологию под брендом Optane. Intel планирует запустить DIMM-вариант 3DXP, Apache Pass, как «большую (но медленную) память» в конце 2018 года;
- Faster NAND (более быстрая NAND) – Samsung и Toshiba создали каждый свой SLC флеш-накопитель как тактический ответ на Optane – ожидаются в продаже в 2018 году. Samsung продвигает свою разработку под брендом «Z-NAND»;
- Resistive RAM (ReRAM, Memristor, резистивная оперативная память) – Toshiba / SANdisk. Технология популяризируется лабораториями HP, но это затрудняет их коммерциализацию. Планы выпуска WD / SNDK перенесены на 2019 год;
- RAM Phase Phase (PCRAM) – SK Hynix. Разработка технологически аналогична 3DXP. Hynix планирует поставлять первые устройства PCRAM в 2019 году.

К 2020 году системные архитекторы и конечные пользователи будут иметь широкий выбор вариантов дизайна в этой области «быстрого хранения».

Хотя вышеупомянутые технологии подходят как устройства блочного хранения, подключенные к шине PCIe, они не имеют требуемых характеристик производительности и долговечности, чтобы быть настоящей памятью класса DRAM. Из оставшихся технологических возможностей два показывают наибольший потенциал для замены DRAM:

- Magnetic RAM (MRAM, магнитная память) – скорость DRAM-класса, но технология не имеет хороших показателей по стоимости / плотности для поддержки крупномасштабных бизнес-кейсов. Поиск ниши во встроенных приложениях. Everspin – самая заметная компания-стартап в этом области;
- Carbon Nanotube RAM (NRAM, углеродная нанотрубка) – скорость DRAM, малая мощность и возможность масштабирования как в 2D, так и 3D для долгосрочного сокращения затрат. Nantero – известный стартап в NRAM. В то время как ожидается, что каждый из основных FAB-модулей памяти будет предоставлять разнообразные продукты SCM, их фундаментальные технологии, как прогнозируется, не смогут обеспечить настоящую PMEM-замену PMEM в областях производительности, мощности, выносливости или масштаба литографии (стоимость). Из оставшихся потенциальных технологий NRAM является наиболее перспективной из-за ее уникальных способностей для решения проблемы замены DRAM и, следовательно, потенциально ускоряет развитие зарождающегося рынка PMEM и окружающей экосистемы оборудования и программного обеспечения.

Когда ожидается выход Persistent Memory на внешний рынок?

Intel готовит своих клиентов к продукту PMEM NVDIMM в конце 2018 года. Эти продукты должны быть быстро приняты OEM-производителями и поддерживаться независимыми поставщиками ПО. В результате усилий Intel по развитию рынка, другие новые технологии памяти, которые будут выходить на рынок, будут более легко им восприниматься, поскольку достигнут коммерциализации в течение следующих 2 лет.

Со временем доля SCM будет возрастать и замещать традиционные уровни хранения на HDD/SSD (рис. 4).

Публикация подготовлена по материалам, предоставленным Dell EMC

Dell EMC ускоряет внедрение ИИ

Август 2018 г. — Компания Dell EMC объявила о доступности готовых решений Ready Solutions for AI с технологиями Hadoop для машинного обучения и NVIDIA для глубинного обучения. Готовые решения Ready Solutions от Dell EMC упрощают среды искусственного интеллекта (ИИ) и помогают быстрее получать более глубокую аналитическую информацию по сравнению с конкурирующими решениями[1]. С помощью новых комплексных предложений Dell EMC помогает организациям полностью реализовывать потенциал новой технологии.

В ближайшее десятилетие новые технологии, такие как ИИ, изменят жизнь людей, а также модели работы и ведения бизнеса. Согласно исследованию (*https://www.delltechnologies.com/ en-us/perspectives/realizing-2030.htm*) Dell Technologies и VansonBourne, в рамках которого были опрошены 3800 руководителей компаний со всего