



Hewlett Packard
Enterprise

Опыт тестирования Optane DCPMM в России под разные задачи

Олег Васьков

Руководитель центра высоких технологий,
Группа корпоративных решений

Oleg.Vaskov@hpe.com

29.10.2019

Программа

1

HPЕ Persistent Memory для HPЕ ProLiant Gen10
Настройка. Область применения

2

Тестирование HPЕ Persistent Memory в России и мире

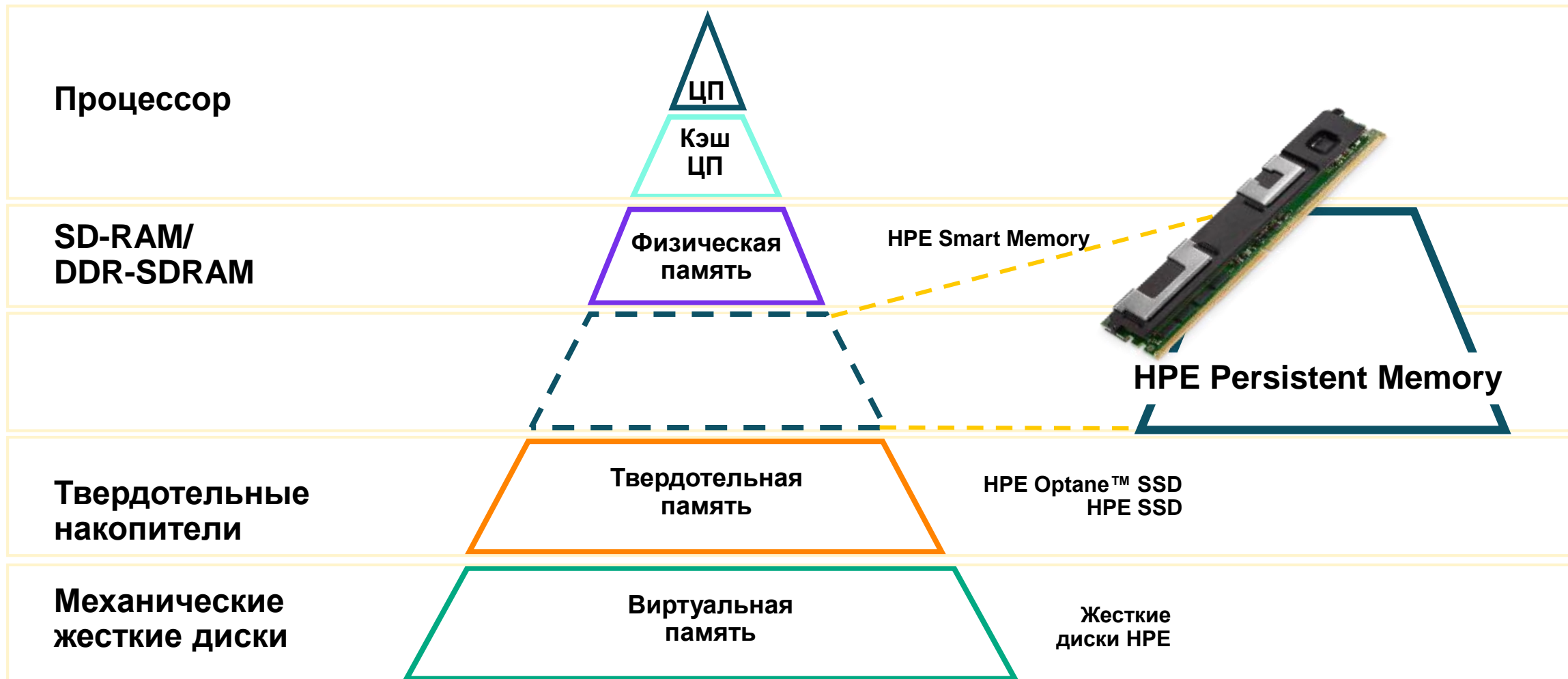
3

Дальнейшие шаги



HPE Persistent Memory

Новый взгляд на иерархию памяти и системы хранения



Портфель HPE Persistent Memory для Gen10

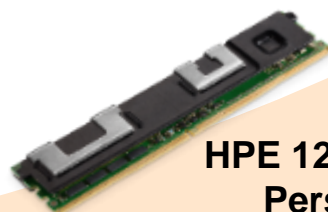
Существенное расширение портфеля памяти класса хранения

Емкость и целевой сегмент/рынок

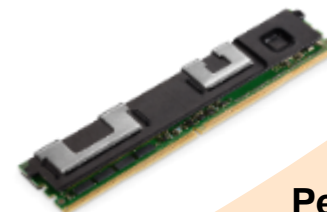
Узкие места в СХД
Оптимизация лицензий на ПО
Кэширование



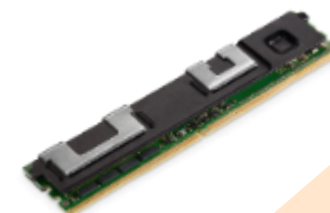
**HPE 16GB NVDIMM
Single Rank x4 DDR4 2666
Module Kit**



**HPE 128GB 2666
Persistent
Memory Kit**



**HPE 256GB 2666
Persistent Memory Kit**



**HPE 512GB 2666
Persistent Memory Kit**

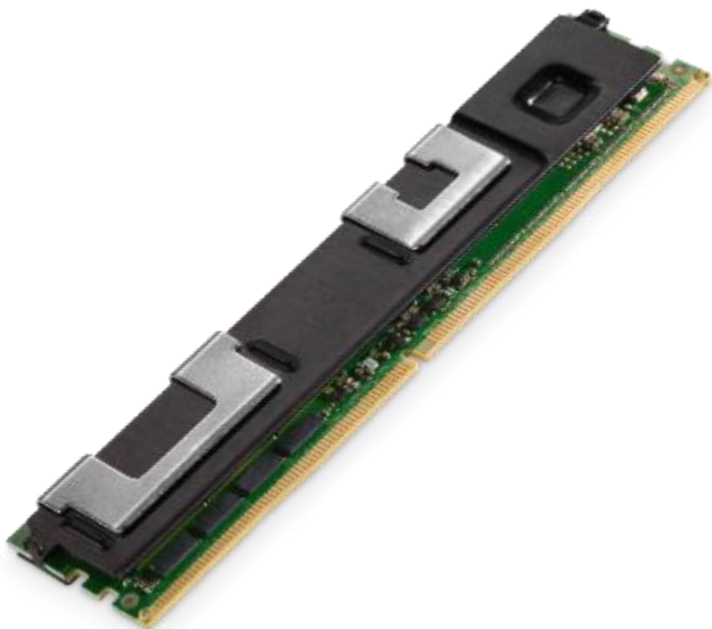
Вычисления в памяти
Контрольные точки с
возможностью восстановления
HTAP и аналитической системы
Большие базы данных
Программно-определяемая СХД
Поставщики услуг,
виртуализация

Емкость и целевой сегмент/рынок



HPE Persistent Memory

Новые технологии для PMEM на базе Intel® Optane DC



Новые модули памяти HPE Persistent Memory доступны в вариантах емкости **128, 256 и 512 Гбайт**, на базе Intel® Optane DC и созданы для обеспечения больших объемов памяти и быстрой работы СХД

Идеальное решение для:

- повышения общей производительности системы и снижения задержек доступа к данным за счет их размещения ближе к процессору на энергонезависимых носителях
- второго поколения процессоров Intel Xeon Scalable

Ключевые характеристики:

- Возможность использования и как в качестве памяти большого объема, так и в качестве высокоскоростной СХД
- Емкость до 3 Тбайт на процессор (2 процессора: 6 Тбайт, 4 процессора: 12 Тбайт)
- Объем в 4 раза больше памяти для серверов DDR4
- Производительность в 3 раза выше, чем у твердотельных накопителей NVMe
- Более высокое соотношение емкости в расчете на доллар, чем при использовании обычных DRAM

Поддержка платформ и операционных систем:

- ProLiant DL360, DL380, DL560, DL580, SDFlex, Synergy 480, 660 и Apollo 2000
- MS Windows Server 2019, RHEL 7.6, SUSE Linux Enterprise Server 12 SP4, SUSE Linux Enterprise Server 15, VMware 6.7 U1

Подробнее: <https://www.hpe.com/info/persistentmemory>



HPE Persistent Memory

Режимы работы PMEM настраиваются BIOS сервера и/или через REST API

Режим Memory Mode Как обычный DRAM

Производительность, сравнимая с производительностью DRAM, с низкими задержками¹



Большой объем по сравнению с использованием обычных DRAM модулей

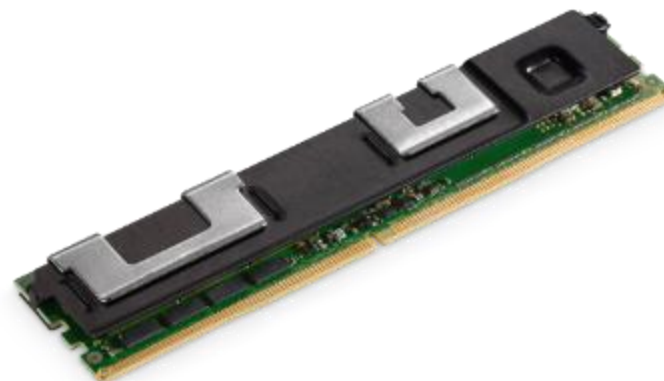


Более низкая стоимость по сравнению со сравнимыми модулями DRAM (цена за Гбайт)



Просто использовать:³работает с современными приложениями

Сочетание возможностей модуля DRAM и СХД в одном продукте



Смешанный режим
Делим память между режимами Memory Mode и App Direct

Режим App Direct Атрибуты СХД

Постоянное хранение данных, объем существенно больше, чем у модуля DRAM²



Надежное хранение: хранение данных даже во время циклов включения-выключения без потребности во внешних источниках питания



Высокая доступность и меньше простоя: ускорение времени на перезагрузку; сокращение времени ожидания при выборке данных



Значительное ускорение работы СХД: перемещение, хранение и обработка более крупных наборов данных, ближе к процессору

1. «производительность, сравнимая с производительностью DRAM»: ожидается, Intel Optane будет работать с задержками сравнимой с DDR4 DRAM. Расширенные результаты тестирования планируется представить позже. «Низкие задержки»: задержки при передаче данных по шине памяти практически незаметны, по сравнению с задержками при передаче данных по шине ввода/вывода PCIe на NAND носитель или жесткий диск. Результаты тестирования и подробные технические детали, планируется предоставить позднее.
2. Энергонезависимая память HPE Persistent Memory имеет 3 разных объема: 128 Гбайт, 256 Гбайт, 512 Гбайт. Традиционные доступные к продаже DIMM модули памяти DDR4 DRAM обладают максимальным объемом 128 Гбайт.
3. Перед использованием памяти HPE DC Persistent Memory потребуется обновление BIOS.

Настройка режима работы Persistent Memory в HPE Gen10

Задаем долю Volatile (“Memory Mode”) и Persistent (“App direct”) в BIOS и/или через REST API

Узел: HPE ProLiant DL380 Gen10 - 1*6240 (2.6GHz/18-core/150W) + **6*32GB RDIMM 2666** + **1*512GB PMEM** (Optane DIMM)

Пример – 0% в Volatile и 100% Persistent:

- ОС увидит 192 ГБ оперативной памяти и одно PMEM устройство на 502 ГБ, которое можно использовать в качестве энергонезависимой памяти
- Лицензия, встроенная в процессор (1ТБ, 2ТБ, 4.5ТБ), ограничивает общий объем DIMM памяти установленной на процессор (т.е. складывается объем всех DIMM модулей) и если он превышен, система выдает ошибку

BIOS/Platform Configuration (RBSU)

HPE ProLiant DL380 Gen10
Server SN: CZ2 [REDACTED]
iLO IPv4: 192.168.77.43
iLO IPv6: FE80::7210: [REDACTED]
User Default: OFF

Goal Configuration Options

Current Goal Configuration:

Volatile Memory:	
DRAM used as PMM Cache	0 GB
DRAM used as Volatile Memory	192 GB
PMM Volatile Size	0 GB
Total System Volatile Memory Size	192 GB

Persistent Memory:

Processor 1 PMM Persistent Size	502 GB
Processor 2 PMM Persistent Size	0 GB
Total System PMM Persistent Size	502 GB

Enter: Select
ESC: Exit
F1: Help
F7: Load Defaults
F10: Save
F12: Save and Exit

Volatile Memory Capacity %: 0%
Persistent Memory Interleaving: Disabled

Apply Goal Configuration
Delete Goal Configuration

Exit | Changes Pending | Reboot Required | F7: Load Defaults | F10: Save | F12: Save and Exit



Настройка Persistent Memory в App Direct для Windows 2019 на серверах HPE

Новый тип устройств. Настройка штатными средствами

Узел: HPE ProLiant DL380 Gen10 - 1*8256 (3.8GHz/4-core/105W) + 1*512GB PMEM (Optane DIMM) + 6*64GB LRDIMM

Шаг 1: настраиваем режим работы PMEM в BIOS сервера

Шаг 2: устанавливаем ОС

Шаг 3: ставим HPE SPP (<http://www.hpe.com/info/spp>)

Шаг 4: все очищаем

Get-PmemDisk | Get-PmemPhysicalDevice | Initialize-PmemPhysicalDevice

Шаг 5: размечаем

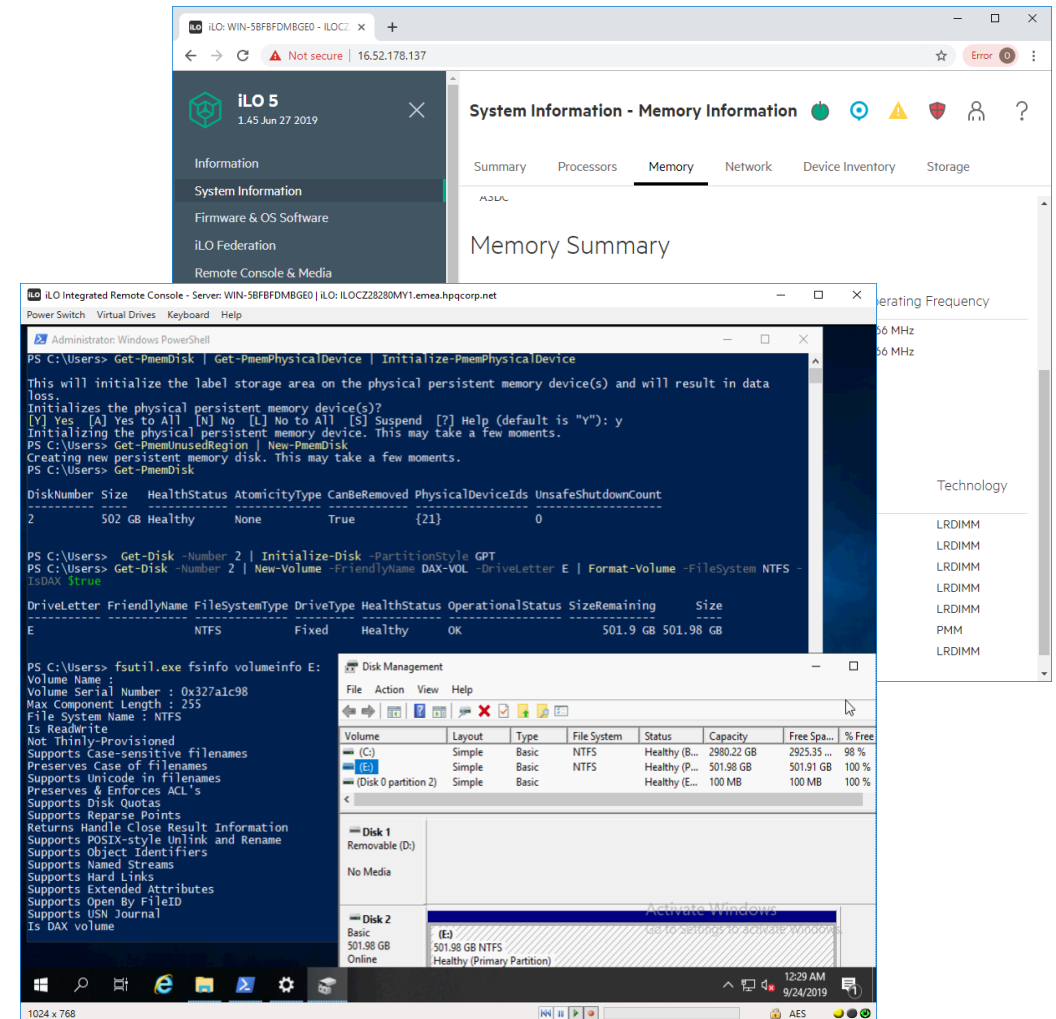
Get-PmemUnusedRegion | New-PmemDisk

Шаг 6: форматируем и монтируем

Get-PmemDisk

Get-Disk -Number 2 | Initialize-Disk -PartitionStyle GPT

Get-Disk -Number 2 | New-Volume -FriendlyName DAX-VOL -DriveLetter E | Format-volume -Filesystem NTFS -IsDAX \$true



Настройка Persistent Memory в App Direct для Linux на серверах HPE

Новый тип устройств. Настройка штатными средствами

Узел: HPE ProLiant DL360 Gen10 - 2*6244 (3.7GHz/8-core/150W) + 192GB (12*16GB RDIMM 2666) + **2*512GB PMEM** (Optane DIMM) + P408i-a/2GB + 2*900GB/10K + 2*240GB SATA + 2*750GB NVMe PCIe

Шаг 1: настраиваем режим работы PMEM в BIOS сервера

Шаг 2: устанавливаем ОС

Шаг 3: ставим HPE SPP (<http://www.hpe.com/info/spp>)

Шаг 4: все очищаем

```
#ndctl disable-namespace all
```

```
#ndctl destroy-namespace all
```

Шаг 5: размечаем

```
#ndctl create-namespace --region=region0 --mode=fsdax #ndctl
```

```
create-namespace --region=region1 --mode=fsdax
```

Шаг 6: форматируем и монтируем

```
#mkfs.xfs -f /dev/pmем0
```

```
#mkdir /mnt/pmем0
```

```
#mount -t xfs -o dax /dev/pmем0 /mnt/pmем0
```

```
#mkfs.xfs -f /dev/pmем1
```

```
#mkdir /mnt/pmем1
```

```
#mount -t xfs -o dax /dev/pmем1 /mnt/pmем1
```

The image shows two screenshots from an HPE iLO 5 interface. The top screenshot displays the 'System Information - Memory Information' page, which includes a table of physical memory slots and their status. The bottom screenshot shows the 'iLO Integrated Remote Console' with a terminal window displaying the execution of various commands and a 'Benchmark' window showing disk performance metrics for a 531 GB Block Device.

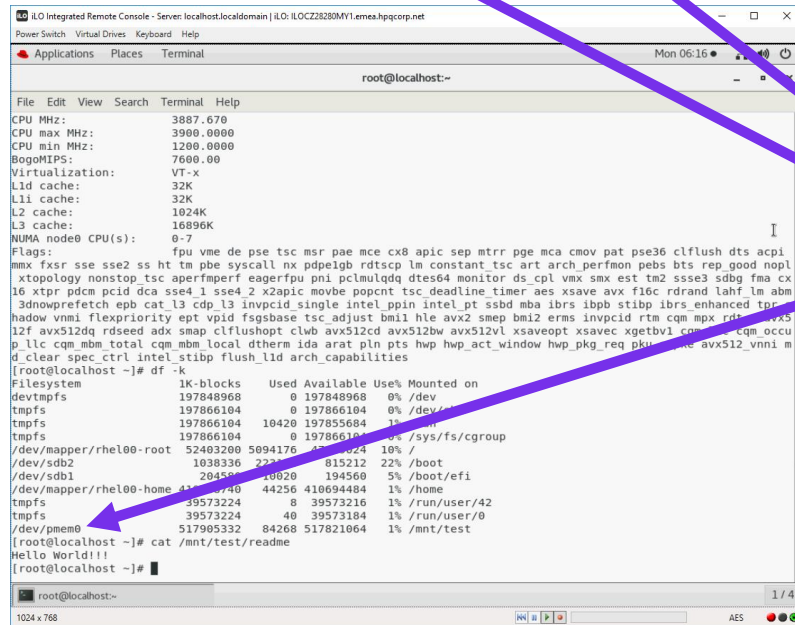
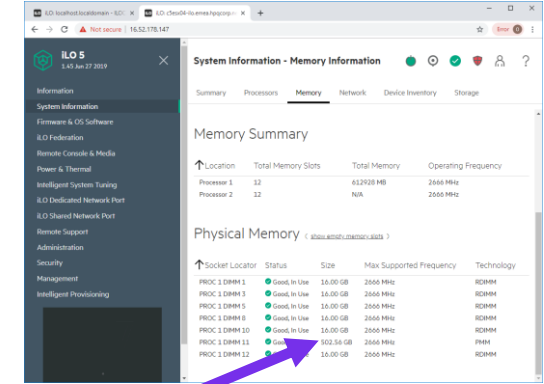
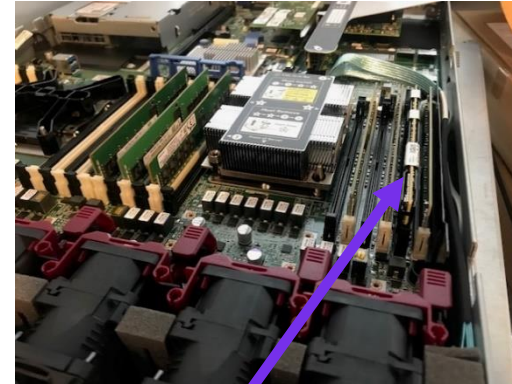
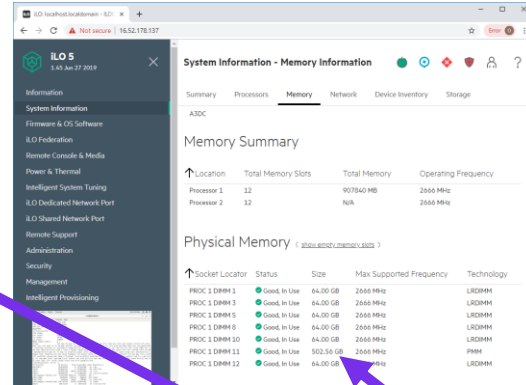
Processor	Core	Memory	Frequency
Processor 1	12	612928 MB	2666 MHz
Processor 2	12	612928 MB	2666 MHz

Socket Locator	Status	Size	Max Supported Frequency	Technology
PROC 1 DIMM 1	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 1 DIMM 3	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 1 DIMM 5	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 1 DIMM 8	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 1 DIMM 10	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 1 DIMM 11	Good, In Use	502.56 GB	2666 MHz	PMEM
PROC 1 DIMM 12	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 1	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 3	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 5	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 8	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 10	Good, In Use	16.00 GB	2666 MHz	RDIMM
PROC 2 DIMM 11	Good, In Use	502.56 GB	2666 MHz	PMEM
PROC 2 DIMM 12	Good, In Use	16.00 GB	2666 MHz	RDIMM

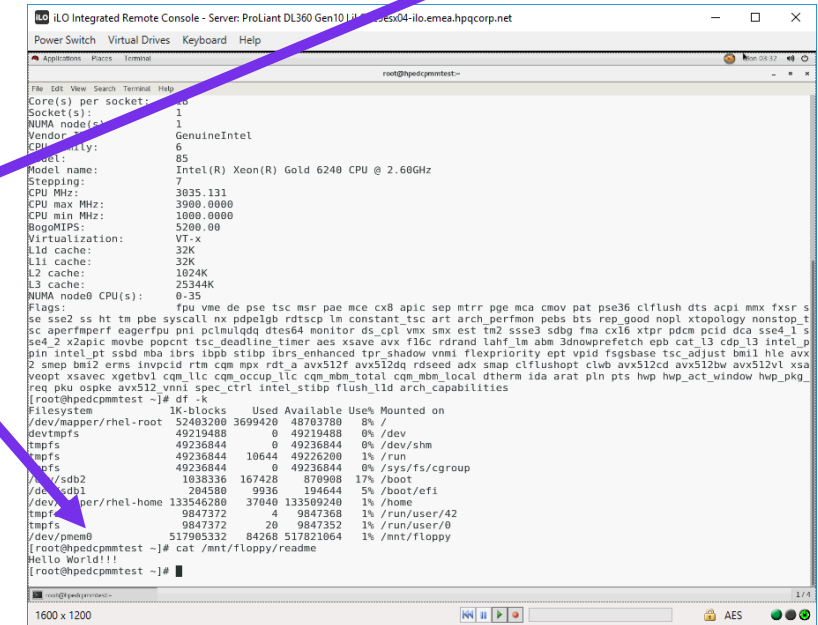
Metric	Value
Average Read Rate	2.9 GB/s (100 samples)
Average Write Rate	1.9 GB/s (100 samples)
Average Access Time	0.00 msec (1000 samples)

Перенос модулей PMEM в режиме Persistent Memory (App Direct)

PMEM DIMM модули могут быть перенесены между серверами, без потери данных



835810-B21
HPE 512GB 2666
Persistent Memory Kit
featuring Intel Optane DC



Программа

1

HPЕ Persistent Memory для HPЕ ProLiant Gen10
Настройка. Область применения

2

Тестирование HPЕ Persistent Memory в России и мире

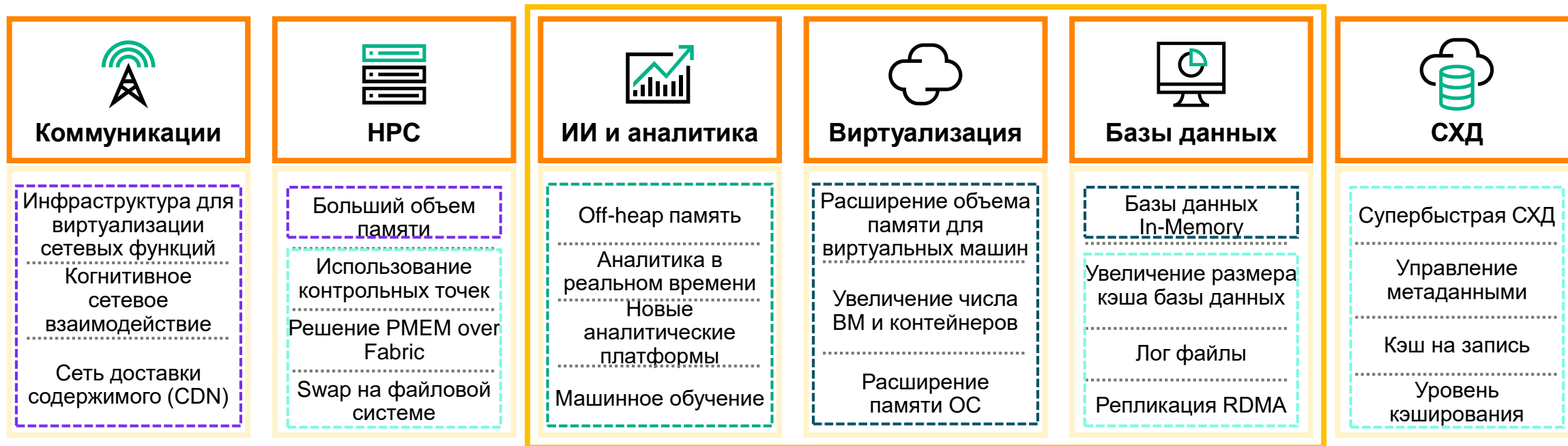
3

Дальнейшие шаги



HPE Persistent Memory

Рабочие нагрузки, сценарии использования и режимы работы



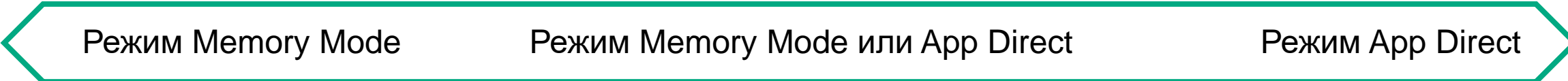
Содержит исполняемые программы и данные, которые процессор использует в данный момент

Сохраняет данные и программы для дальнейшего использования



HPE Persistent Memory

Рабочие нагрузки, сценарии использования и режимы работы



Содержит исполняемые программы и данные, которые процессор использует в данный момент

Сохраняет данные и программы для дальнейшего использования



Persistent Memory в режиме Memory Mode на HPE ProLiant Gen10

VMmark 3.1 – тест виртуализированных приложений

Тест: Vmmark 3.1, 2 Tiles

Тип нагрузки: множество VM базы данных и веб-серверов приложений.

Всего - 38 виртуальных машин, **700ГБ RAM**, 160 vCPU, 3.8ТБ на внешней СХД

Параметры VMmark были модифицированы для увеличения объема памяти

Аппаратная платформа Gen10:

Узел 1: HPE Synergy 480 Gen10 - 2*8260Y (2.4GHz/24-core)
24*64GB LRDIMM 2666

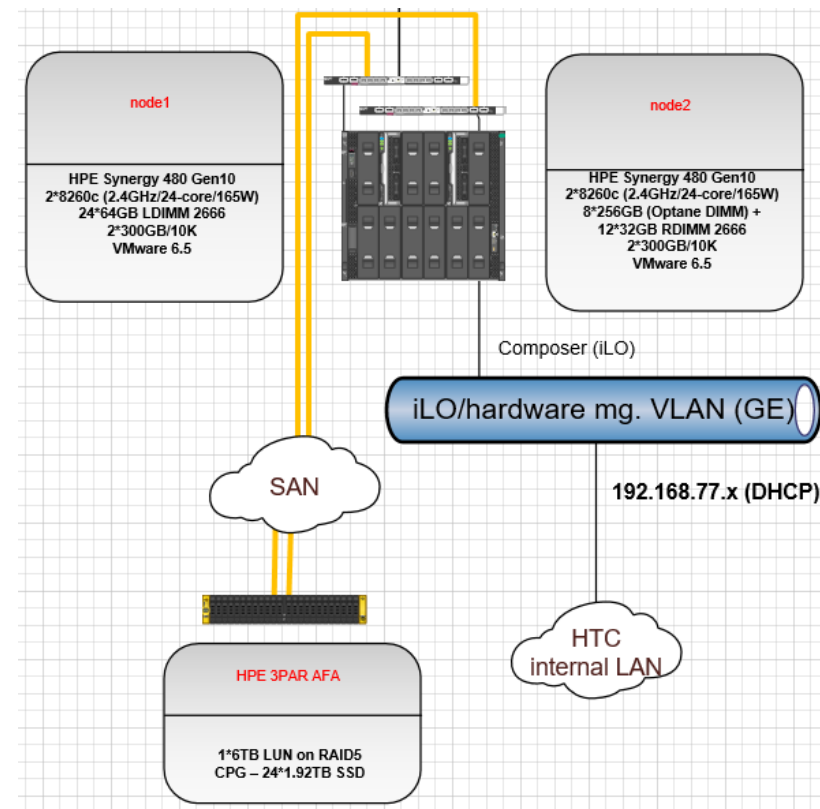
Узел 2: HPE Synergy 480 Gen10 - 2*8260Y (2.4GHz/24-core)
8*256GB (Optane DIMM) + 12*32GB RDIMM 2666*

Режим работы Intel Optane DIMM: **Memory Mode**

Продолжительность теста: 5 часов

Результат:

LRDIMM 2666	Optane + RDIMM* 2666
100%	103%



* Использовались только в качестве кэша для Optane DIMM (т.е. ОС их не видела), т.е. в ESXi видна как 1510 ГБ

** чем выше, тем лучше

*** Для Intel Optane урезали память программно

Persistent Memory в режиме Memory Mode на HPE ProLiant Gen10

VMmark 3.1 – тест виртуализированных приложений

Тест: Vmmark 3.1, 2 Tiles

Тип нагрузки: множество VM базы данных и веб-серверов приложений.

Всего - 38 виртуальных машин, **1000ГБ RAM**, 160 vCPU, 3.8ТБ на внешней СХД

Параметры VMmark были модифицированы для увеличения объема памяти

Аппаратная платформа Gen10:

Узел 1: HPE Synergy 480 Gen10 - 2*8260Y (2.4GHz/24-core)
24*64GB LRDIMM 2666

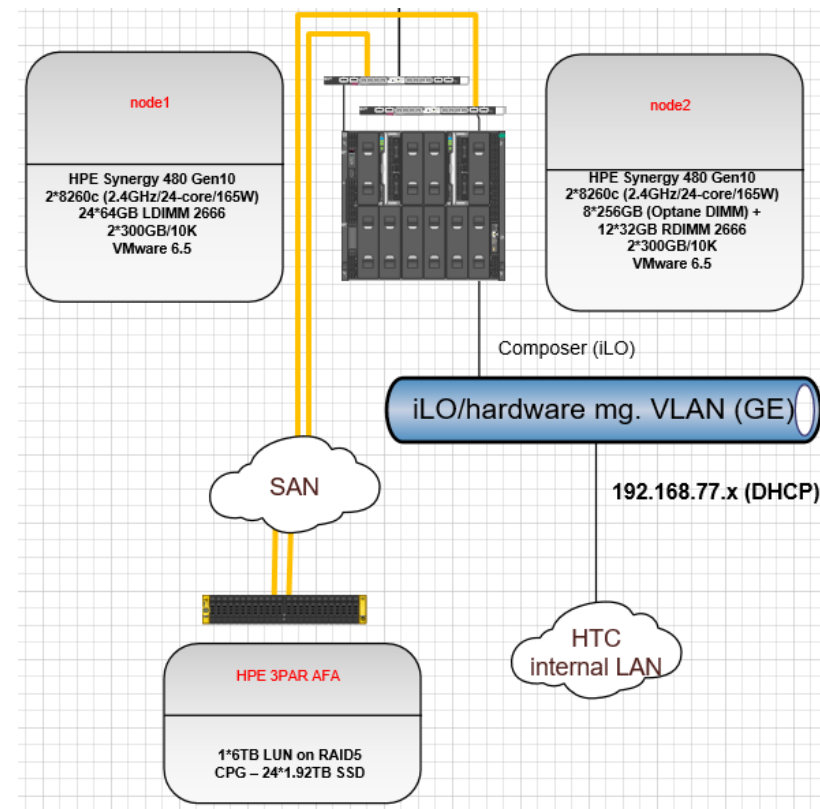
Узел 2: HPE Synergy 480 Gen10 - 2*8260Y (2.4GHz/24-core)
8*256GB (Optane DIMM) + 12*32GB RDIMM 2666*

Режим работы Intel Optane DIMM: **Memory Mode**

Продолжительность теста: 5 часов

Результат:

LRDIMM 2666	Optane + RDIMM* 2666
100%	97%



* Использовались только в качестве кэша для Optane DIMM (т.е. ОС их не видела), т.е. в ESXi видна как 1510 ГБ

** чем выше, тем лучше

*** Для Intel Optane урезали память программно

HPE Persistent Memory

Рабочие нагрузки, сценарии использования и режимы работы



Режим Memory Mode

Режим Memory Mode или App Direct

Режим App Direct

Содержит исполняемые программы и данные, которые процессор использует в данный момент

Сохраняет данные и программы для дальнейшего использования

Самая большая отдача

Режим Memory Mode

Режим Memory Mode или App Direct

Режим App Direct

Persistent Memory в режиме App Direct на HPE ProLiant Gen10

vdbench на сырых устройствах

Тест: vdbench*

Аппаратная платформа Gen10:

HPE ProLiant DL380 Gen10 – 1*8256 (3.8GHz/4-core/105W) + 1*512GB PMEM (Optane DIMM) + 6*64GB LRDIMM

Режим работы PMEM (Optane DIMM): **Persistent (App Direct)**

Программный стек

Red Hat Enterprise Linux 7.6



Типы энергонезависимой памяти

- 1) Smart Array P408i-a RAID1 из 2*600GB/10K SAS - /dev/sdc
- 2) Smart Array P408i-a RAID1 из 2*1.92TB SATA SSD (Intel S4510) - /dev/sdc
- 3) Smart Array P408i-a RAID1 из 2*480GB SATA SSD (Intel S4600) - /dev/sdd
- 4) NVMe – 3.2TB (Intel P4600) - /dev/nvme1n1
- 5) NVMe – 750GB (Intel Optane P4800X) – /dev/nvme0n1
- 6) PMEM – 512GB (Intel Optane DIMM) – /dev/pmem0

Результат:

#	Test name	i/o rate	MB/sec	bytes i/o	read pct	resp time	read resp	write resp	read max	write max	resp stddev	queue deps	cpu% sys+u	cpu% sys
1	HPE Smart Array P408i-a RAID1 2*600GB/10K SAS	313	8	25 327	59,4	12,759	21,451	0,040	184,28	0,15	19,590	4,0	5,0	4,5
2	HPE Smart Array P408i-a RAID1 2*1.92TB SATA SSD (Intel S4510)	11 657	278	25 023	59,9	0,339	0,377	0,283	5,64	4,58	0,443	4,0	2,2	1,2
3	HPE Smart Array P408i-a RAID1 2*480GB SATA SSD (Intel S4600)	18 341	435	24 895	60,0	0,215	0,191	0,251	5,23	5,14	0,261	3,9	3,2	1,8
4	PCI NVMe – 3.2TB (Intel P4600)	18 587	442	24 907	60,0	0,212	0,34	0,022	5,51	0,34	0,427	3,9	2,0	0,7
5	PCI NVMe – 750GB (Intel Optane P4800X)	83 991	1 988	24 821	59,9	0,045	0,043	0,047	4,79	1,13	0,050	3,8	8,2	4,1
6	PMEM – 512GB (Intel Optane DIMM) - 1*512GB	170 185	4 037	24 871	60,0	0,022	0,021	0,024	0,57	0,95	0,054	3,8	51,0	47,2

*sd=sd1,lun=/dev/nvme1n1,openflags=o_direct
 wd=wd060,sd=sd*,xfersize=(4k,48,8k,19,16k,5,32k,9,48k,13,64k,3,128K,1,256k,1,512k,1),rdpct=60,seekpct=70,rhpct=0
 rd=run060,wd=wd060,iorate=max,elapsed=10,threads=4

Схема тестового стенда HPE и описание теста

Postgres Pro 11.2*

- **Аппаратная платформа**

Сервер: HPE Synergy 660 Gen10 - 4*8260Y (2.4GHz/24-core) + 8*256GB PMEM (Optane DIMM) + 24*64GB LRDIMM + Smart Array P416ie-m + 4*1.92TB SATA SSD RAID10 + CNA 3820C (с функцией 8Gb FC)

- Режим работы PMEM (Optane DIMM): **Persistent (App Direct)**

- СХД: HPE 3PAR 7450c - 24*1.92TB SSD RAID5

- **Программный стек**

Red Hat Enterprise Linux 7.6
СУБД Postgres Pro Enterprise 11.2

- **Тест**

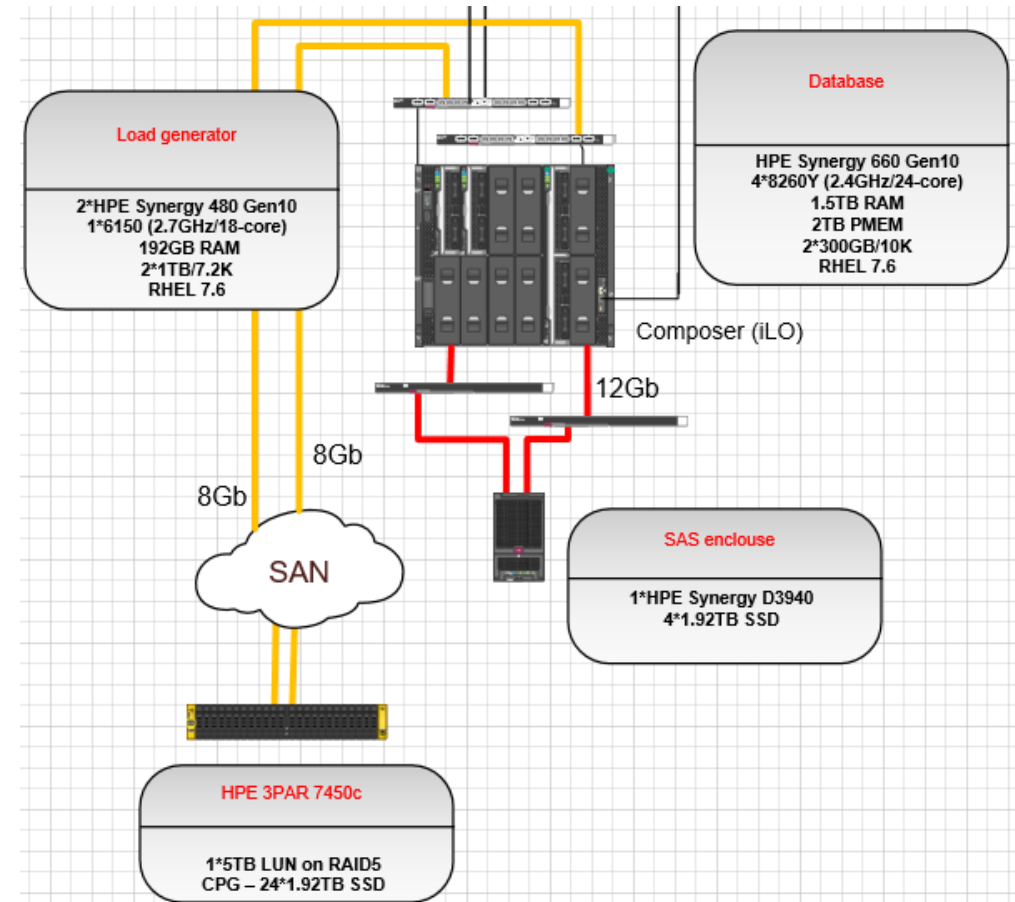
синтетический тест на транзакционную нагрузку, смесь INSERT/UPDATE (очень похож на TPC-B)

- **Типы энергонезависимой памяти**

Smart Array – 2ТБ – RAID10 из 4*1.92TB SATA SSD - /mnt/smartarray

PMEM – 512ГБ – RAID0 из 2*256GB – /mnt/pmем

3PAR – 5.5ТБ – RAID5 на дисковой группе 24*1.92TB SSD – /mnt/3par



* Построена на базе PostgreSQL 10.x.x и существенно дополнена дополнительными пакетами, включая пакеты разработанные Postgres Professional

Результат тестирования Postgres Pro на HPE Synergy Gen10

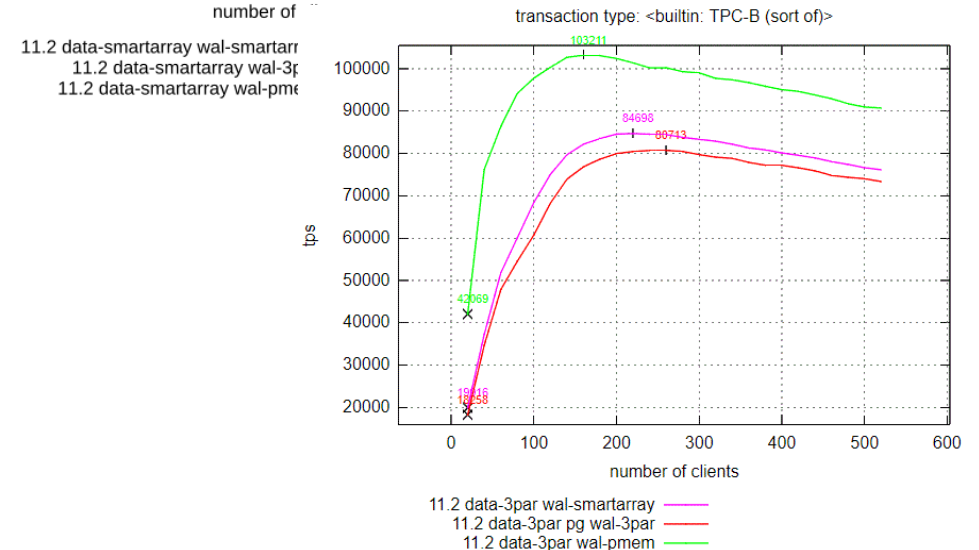
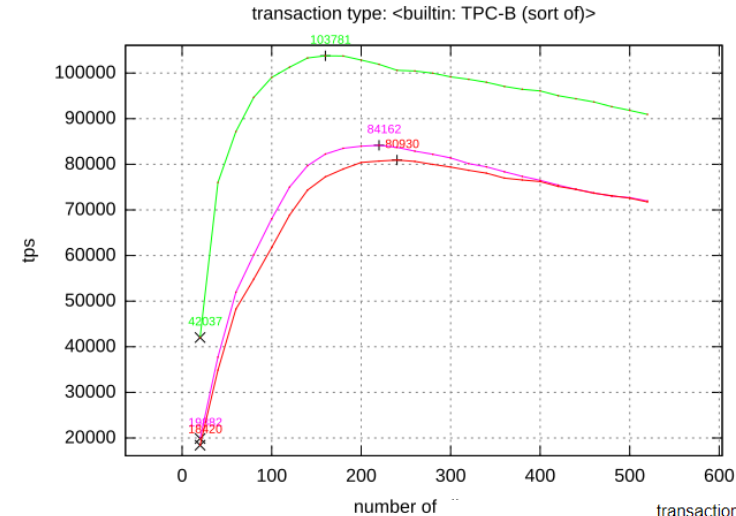
HPE Persistent Memory для хранения лог файлов повышает производительность

• Результат

Perf. (tps)	Database	Log	%
102 000	PMEM	PMEM	22%
102 000	3PAR	PMEM	22%
80 000	3PAR	3PAR	0%

Perf. (tps)	Database	Log	%
102 000	PMEM	PMEM	22%
102 000	Smart Array	PMEM	22%
83 900	Smart Array	Smart Array	0%

- Размер базы – **15ГБ**
Число одновременных активных клиентов - 200
- **Выводы**
 - Использование PMEM для хранения логов базы данных позволяет получить существенный прирост производительности
 - В текущем ядре Linux использовать PMEM следует без программных RAID (т.е. без LVM и/или MD)



Результат тестирования Postgres Pro на HPE Synergy Gen10

HPE Persistent Memory для хранения лог файлов повышает производительность

• Результат

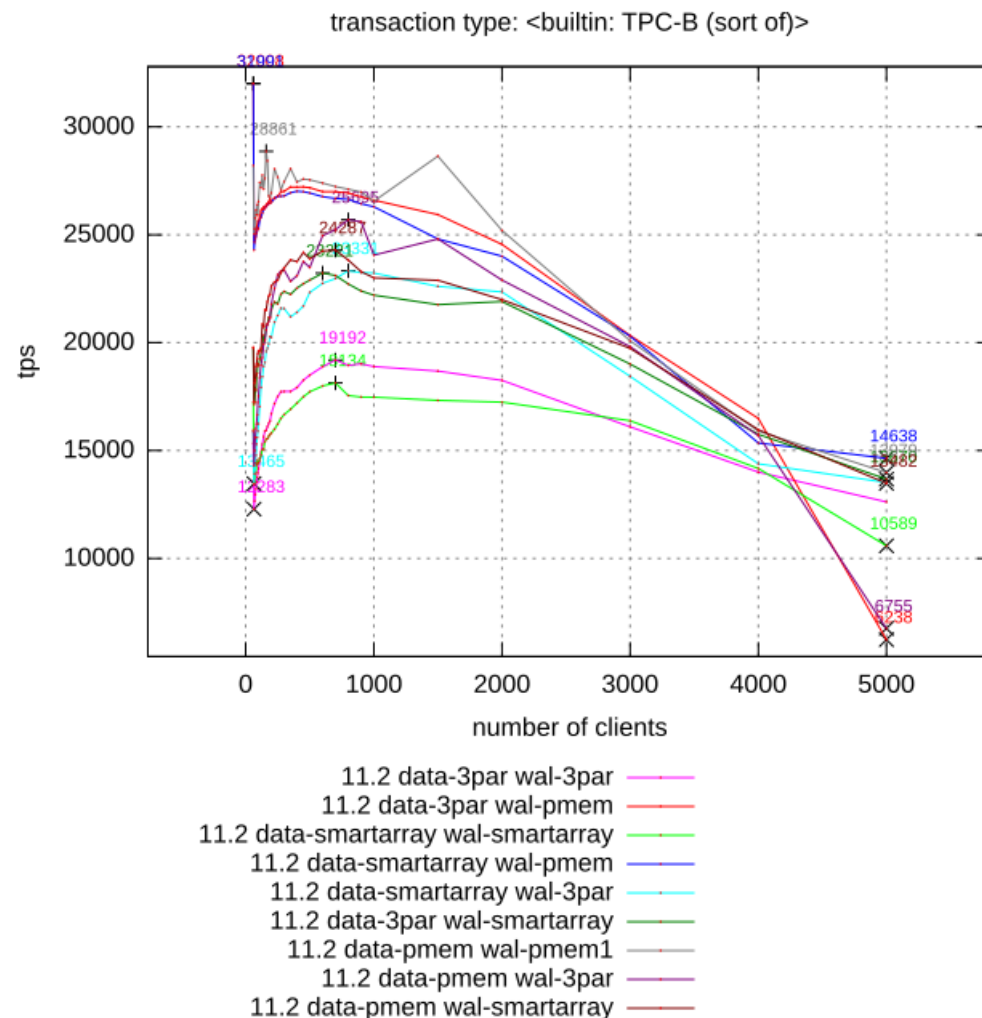
Perf. (tps)	Database	Log	%
25 195	PMEM	PMEM	38%
24 550	3PAR	PMEM	35%
18 248	3PAR	3PAR	0%

Perf. (tps)	Database	Log	%
25 195	PMEM	PMEM	46%
24 001	Smart Array	PMEM	39%
17 243	Smart Array	Smart Array	0%

- Размер базы – **525ГБ**
Число одновременных активных клиентов - 2000

• Выводы

- Использование PMEM для хранения логов базы данных позволяет получить существенный прирост производительности
- В текущем ядре Linux использовать PMEM следует без программных RAID (т.е. без LVM и/или MD)



HPE Persistent Memory

Повышаем бизнес-показатели и улучшаем совокупную стоимость владения СУБД Oracle



96%

производительности от 4-х процессорной системы с обычной DRAM памятью, на 2-х процессорном сервере с PMEM памятью

14%

Ускорение транзакций при использовании PMEM для журнала повтора (redo-log)

Финансовый сектор

Розничная торговля

Образовательные услуги

Цепочки поставок

ORACLE®

Технические подробности тут: <https://h20195.www2.hpe.com/v2/GetDocument.aspx?docname=a00074526enw>

Технические подробности тут: <https://h20195.www2.hpe.com/v2/GetDocument.aspx?docname=a00074230enw>

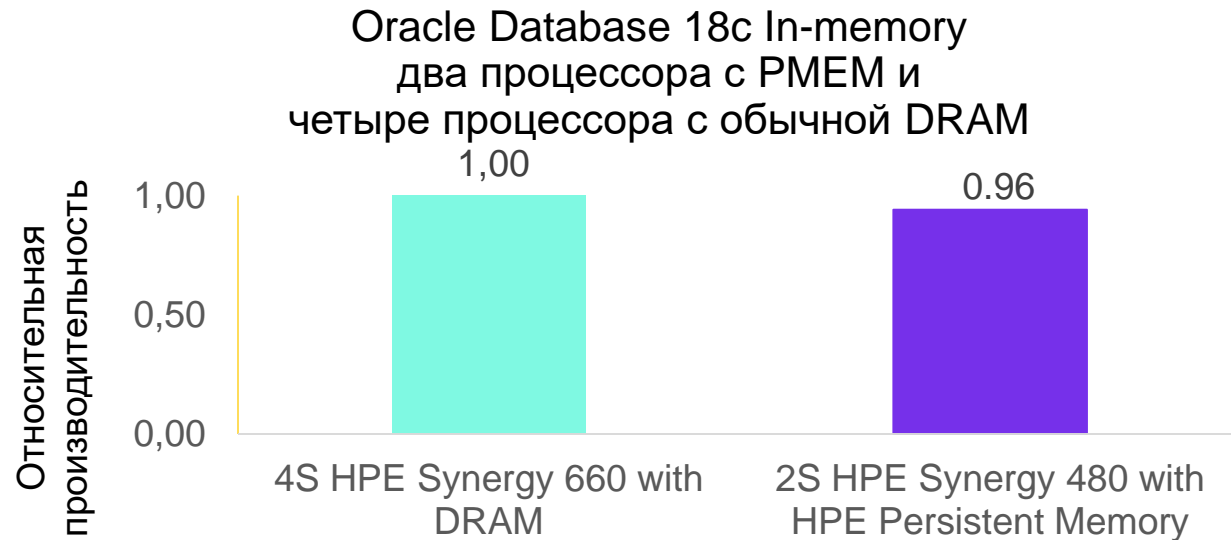


HPE Persistent Memory

СУБД Oracle и вычисления в памяти для бизнес аналитики и OLAP

Настройка/конфигурация тестового стенда

- Тип тестовой нагрузки – OLAP
 - 22 запроса со степенью параллелизма 56
 - Размер базы – 2 ТБ
- Сравнимы конфигурации
 - HPE Synergy 660 Gen10
 - 4*8280 (2.7GHz/28-core/205W)
 - 48*64 ГБ памяти DRAM
 - HPE Synergy 480 Gen10
 - 2*8280M (2.7GHz/28-core/205W)
 - 12*256 ГБ памяти PMEM
 - 12*64 ГБ памяти DRAM (используется как кэш для PMEM)
- СУБД Oracle Database 18c
- Режим работы PMEM – Memory Mode (Volatil_



Результат

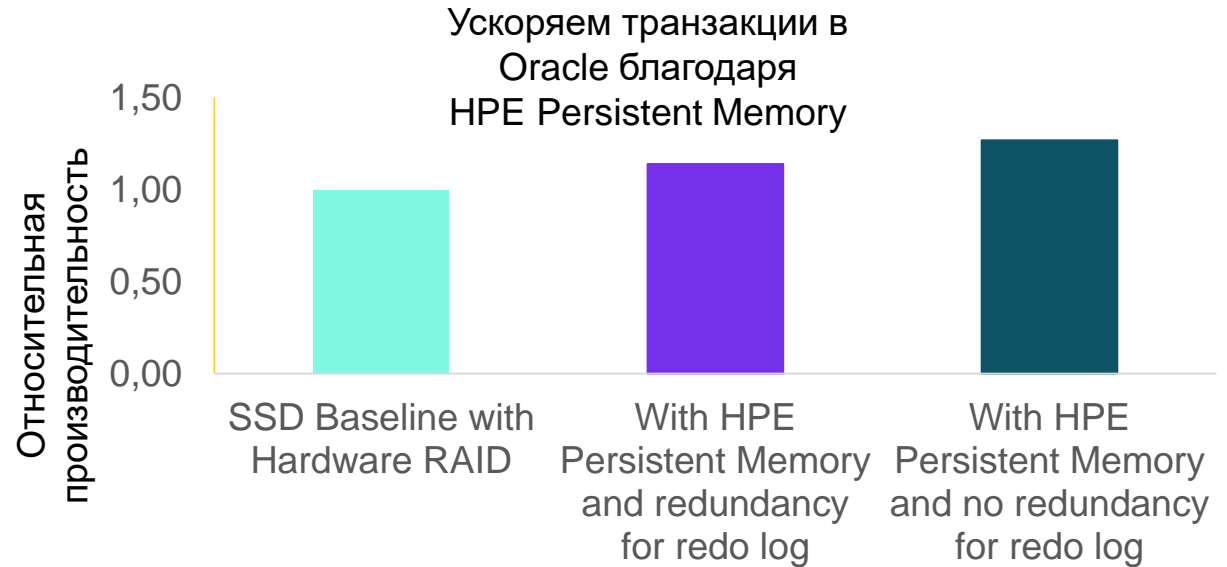
- Двухпроцессорный сервер с HPE Persistent Memory *показал сравнимую производительность, что и четырехпроцессорный с обычными DRAM модулями памяти, а лицензий требует в два раза меньше*

HPE Persistent Memory

Ускорение журнала повтора (redo-log) в Oracle для транзакционных нагрузок (OLTP)

Настройка/конфигурация тестового стенда

- Тип тестовой нагрузки – OLTP
- Сервер HPE Synergy 660 Gen10
 - 4*8280 (2.7GHz/28-core/205W)
- Дисковая полка HPE Synergy D3940
- СУБД Oracle Database 18c
- Сравнимые конфигурации для хранения журнала повтора (redo-log)
 - SmartArray на дисках 400 ГБ SSD WI
 - 4*128 ГБ памяти PMEM
- Режим работы PMEM – App Direct (Persistent)



Результат

- Использование HPE Persistent Memory в режиме App Direct для хранения журналов повторов Oracle ускорило производительность транзакции на 27%, без резервирования журнала. При резервировании журнала, производительность выросла на 14%



HPE Persistent Memory

Улучшает бизнес-результаты ускоряя работу систем принятия решений для Spark SQL



До **26** раз
быстрее принятие
ключевых для
руководителей
решений

В **8,4** раза
выше
производительность

Панель индикаторов

Медицинская диагностика

Сельское хозяйство

Транспортные системы

 Spark SQL

HPE Persistent Memory

Аналитика больших данных с Spark

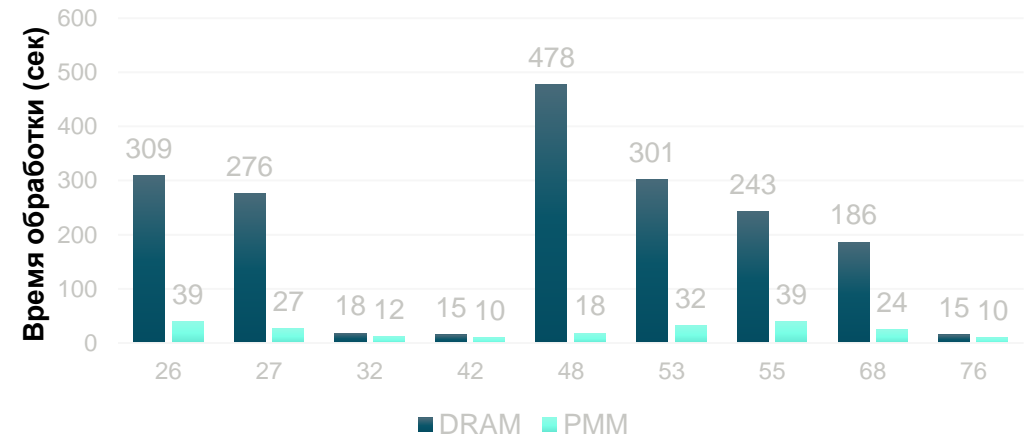
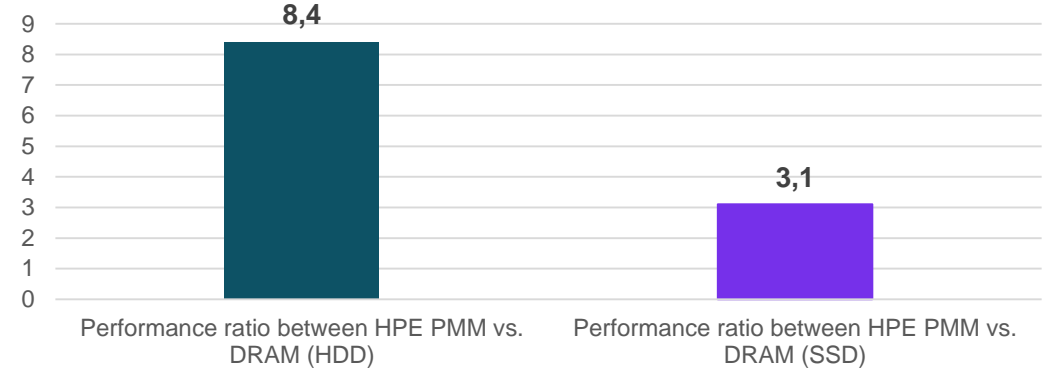
Настройка/конфигурация тестового стенда

- Девять тестов из набора TPC-DS с большим вводом-выводом
- Сравнимые конфигурации
 - память 768 ГБ (24*32 ГБ DRAM)
 - память 192 ГБ (12*16 ГБ DRAM) + 1 ТБ PMEM (12*128 ГБ)
- Режим работы PMEM – **App Direct (Persistent)**
- Установка пакета аналитики для Spark
 - 1 ТБ PMEM энергонезависимой памяти сравниваем с 610 ГБ обычной памятью DRAM
 - Обновление настроек кеша Spark OAP Cache для использования PMEM, а не оперативной памяти

Результат

- Повышение производительности до 26 раз для некоторых тестов
- Общая производительность выросла в 8 раз для конфигураций с жесткими дисками
- Общая производительность выросла в 3 раза для конфигураций с твердотельными дисками

Прирост производительности для Spark



HPE Persistent Memory

Существенный прирост скорости проведения фармацевтических исследований



До **21** %
быстрее
выполняется поиск
коэффициентов,
определяющих
химическое сходство

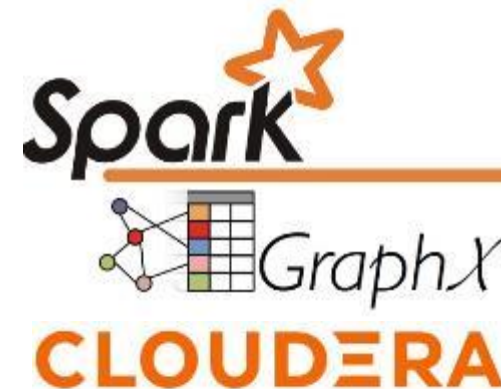
На **48** %
лучше соотношение
цена/
производительность

Умные города

Страхование

Предотвращение
мошенничества

Фармацевтика



Технические подробности тут: <https://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00074594enw>

HPE Persistent Memory

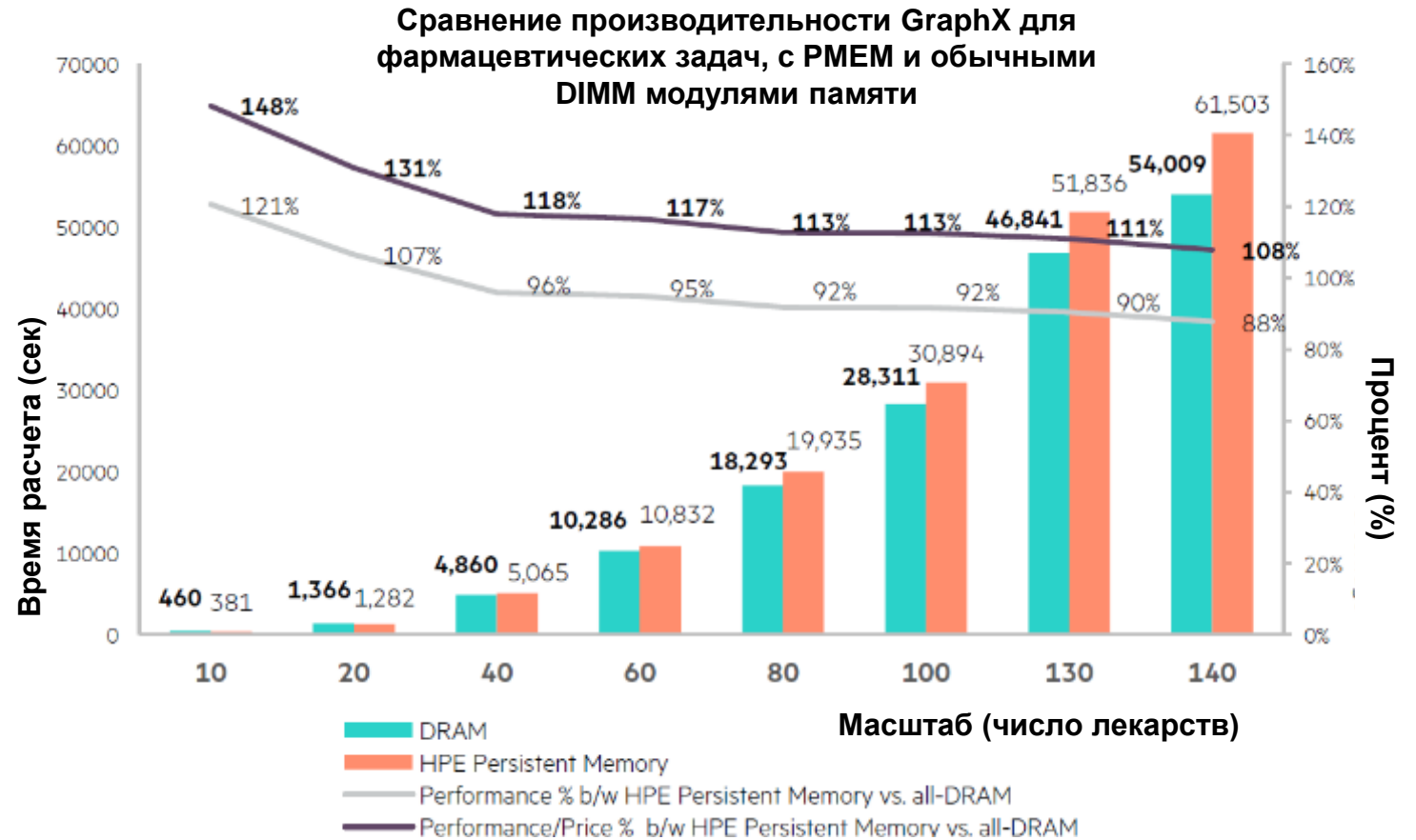
GraphX на аналитической платформе Cloudera

Преимущества HPE Persistent Memory

- Хорошее улучшение соотношения цена/производительность при использовании HPE Persistent Memory для фармацевтических задач с GraphX в сравнении с использованием обычных DRAM модулей памяти

Сравнение производительности

- HPE Persistent Memory общим объемом 1,5 ТБ (DIMM модулями объемом 128 ГБ) показали улучшение соотношения цена/производительности на 48 % лучше, обычные DDR4 DRAM модули с аналогичной емкости
- Большой объем PMEM позволяет увеличить масштаб, не увеличивая число серверных узлов



HPE Persistent Memory

GraphX на аналитической платформе Cloudera

Настройка/конфигурация тестового стенда

- Задача - поиск коэффициентов химического сходства
- Сравниваем разные химические медицинские препараты
- Сравниваемые конфигурации
 - память 1,5 ТБ (24*64 ГБ DRAM)
 - память 1,5 ТБ (с соотношением память/кэш – 1:4)
 - 12*128 ГБ памяти PMEM
 - 12*32 ГБ памяти DRAM (используется как кэш для PMEM)
- Режим работы PMEM – **Memory Mode (Volatile)**
- Четыре вычислительных узла и три узла управления

Результат

- HPE Persistent Memory с модулями объемом 128 ГБ показали соотношение цена/производительность на 48% лучше, чем обычная память DDR4 с аналогичного размера
- Использование PMEM модулей большего объема позволит увеличить масштаб, без увеличивая число серверных узлов

CLLOUDERA



Тестирование 7



HPE Gen10 с PMEM для NoSQL баз данных

Качественный прирост производительности IoT и больших данных

▪ CPU

- 2 x 28-core **Cascade Lake** Procs
 - CPU 0000% @ (fam: 06, model: 55, stepping: 05)
 - CLX SP 28c 2.5GHz 205W
 - L1 = 1792 KB, L2 = 28672 KB, L3 = 39424 KB

▪ Storage

- NVMe Controller
 - 3 x 1600GB SSD
- HPE Smart Array P408i-a SR Gen10
 - 2 x 400 GB SSD
 - 1 x 480 GB SSD

▪ Memory

- 2.6 GHz
- Total = 1.75 TB (2x12 Slots)
 - RDIMM = 16.00 GB x 6
 - 96 GB * 2 = 192 GB Total
 - **DCPMM** = 126.38 GB x 6
 - 758.16 * 2 = 1516.32 GB



HPE Proliant DL380 Gen10 Server



HPE Persistent Memory for 2nd generation Intel® Xeon® Scalable processors



▪ Network

- Adapter 1 / LOM
 - HPE Eth 10/25Gb 2p 640FLR-SFP28
- Adapter 2 / 1GbE
 - HPE Ethernet 1Gb 4-port 331i Adapter
- Adapter 3 / 100GbE
 - HPE InfiniBand EDR/Ethernet 100Gb 2-port
- Adapter 4 / 10GbE
 - HPE Eth 10/25Gb 2p 621SFP28

“PMEM-aware” NoSQL базы данных

Поддерживают режим App Direct (Persistent)

▪ DEVDAX

- Cassandra (Java) https://github.com/shyla226/cassandra/tree/13981_llpl_engine

▪ FSDAX

- Aerospike Enterprise Server (C) **Version 4.5.0.5-1**
- RocksDB (C++) <https://github.com/pmem/rocksdb>
- Redis (C) <https://github.com/pmem/pmem-redis>
- *Memcached (C)*
<https://github.com/lenovo/memcached-pmem>
- *MongoDB (C++)*
<https://github.com/pmem/pmse>

“PMEM-aware” NoSQL базы данных

Поддерживают режим App Direct

▪ Aerospike

```
./bin/ycsb run \
aerospike -s \
-threads 112 \
-threads 112 \
-P workloads/workloada-bench \
-P as.host=10.20.100.65 \
-P \ as.user=admin \
-P as.user=admin \
-P as.namespace=ycsb \
-P target=250000 \
-P maxexecutiontime=14400 \
> \ outputs/workloada-
bench_4hr-
run_aerospike_500_150.out \
2> outputs/workloada-
bench_4hr-
run_aerospike_500_150.er
r
```

***Reduced threads to 112 (1/cpu) to decrease the excessive server load 200 threads caused**

▪ RocksDB

```
./bin/ycsb run \
rocksdb -s \
-threads 10 \
-P workloads/workloada-bench \
-p target=80000 \
-p maxexecutiontime=14400 \
-P rocksdb.dir=/mnt/mem/ycsb-
rocksdb-data \
> outputs/workloada-
bench_4hr-
run_rocksdb_500_10.out \
2> outputs/workloada-
bench_4hr-
run_rocks_500_10.err
```

***Total threads could not exceed ~10, else ops/sec would decrease considerably**

<https://www.aerospike.com/resources/videos/summit19/ty-hpe/>

▪ Cassandra

```
./bin/ycsb run \
cassandra2-cql -s \
-threads 8 \
-threads 8 \
-P workloads/workloada-bench \
-p target=92000 \
-P hosts=10.20.100.66 \
-P \ user=cassandra \
-P password=cassandra \
-P \ as.namespace=ycsb \
-p maxexecutiontime=14400 \
>
outputs/workloada_run_cassandr
a_4hr_500m-8t-15k.out \
2>
outputs/workloada_run_cassandr
a_4hr_500m-8t-15k.err
```

***4 Hour Mark Reached, job automatically killed, only 251M of 500M Records Processed**

▪ Redis

```
./bin/ycsb run \
redis -s \
-threads 10 \
-P workloads/workloada-bench \
-P redis.host=10.20.100.67 \
-P redis.port=6379 \
-P target=10000 \
-P maxexecutiontime=14400 \
> outputs/workloada-
bench_4hr- run_redis_50_10.out \
2> outputs/workloada-
bench_4hr- run_redis_50_10.err
```

***Would have taken >2.5 days to load 500M records, reduced YCSB benchmark test to 50M records for Redis (only)**

“PMEM-aware” NoSQL базы данных

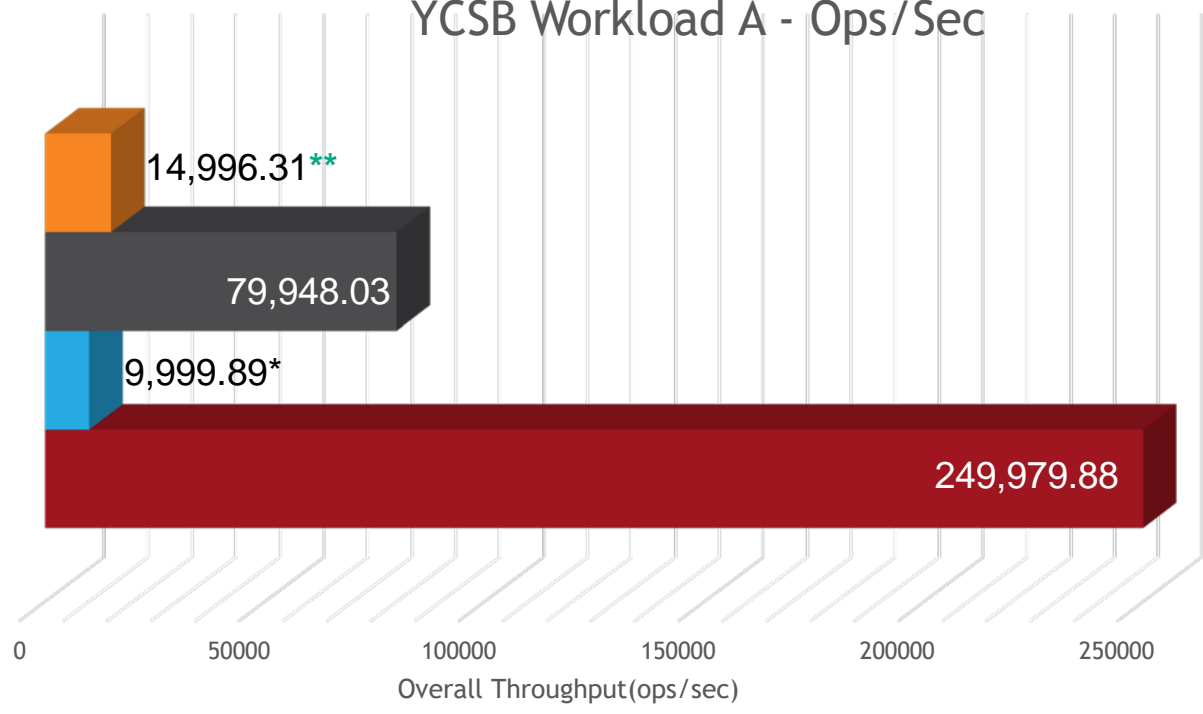
Поддерживают режим App Direct

500 Million 1 Kilobyte Records Processed (50r|50w)

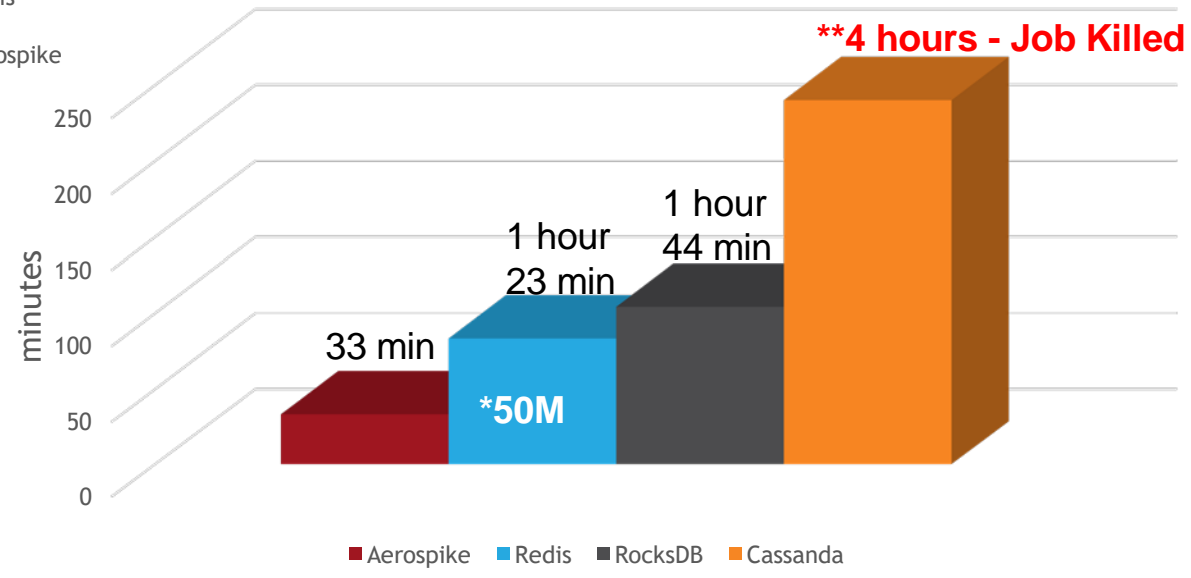
**Redis could not 'load' 500M records in the required timeframe so 50M records was used for its test*

***Cassandra performance dropped 'off a cliff' on reads, writes could be sustained at >90K ops/sec*

YCSB Workload A - Ops/Sec



Overall Runtime - YCSB Workload A



Программа

1

HPE Persistent Memory для HPE ProLiant Gen10
Настройка. Область применения

2

Тестирование HPE Persistent Memory в России и мире

3

Дальнейшие шаги



Практические рекомендации



1

Рассматривай возможность использования PMEM (Optane DCPMM) в режиме Volatile (Memory Mode) для серверов, с оперативной памятью свыше 512ГБ на процессор

2

Исследую возможность использования PMEM (Optane DCPMM) в режиме Persistent (App Direct) для традиционных серверов баз данных

3

Обязательно использую PMEM (Optane DCPMM) в Persistent (App Direct) для NoSQL баз данных*

*требуется современная/свежая версия базы данных с поддержкой PMEM

Спасибо



Будьте в курсе событий HPE

- Слушайте [вебинары](https://gateway.on24.com/wcc/gateway/elitehpeemeademandgen/1474757) о новинках HPE (https://gateway.on24.com/wcc/gateway/elitehpeemeademandgen/1474757)
- Читайте новости и обзоры на [Хабре](https://habrahabr.ru/company/hpe/) (https://habrahabr.ru/company/hpe/)
- Следите за сообществом [HPE в России](#)
- [YouTube](https://www.youtube.com/playlist?list=PL6gp3wcNi2bBK9CuNLtDY0WGLrJIGsu8a) канал HPE (https://www.youtube.com/playlist?list=PL6gp3wcNi2bBK9CuNLtDY0WGLrJIGsu8a)

Подписывайтесь на каналы в социальных медиа:

- **Twitter:** [@HPE_RU](#)
- **Вконтакте:** vk.com/hperussia
- **Telegram:** t.me/hpedigitize

