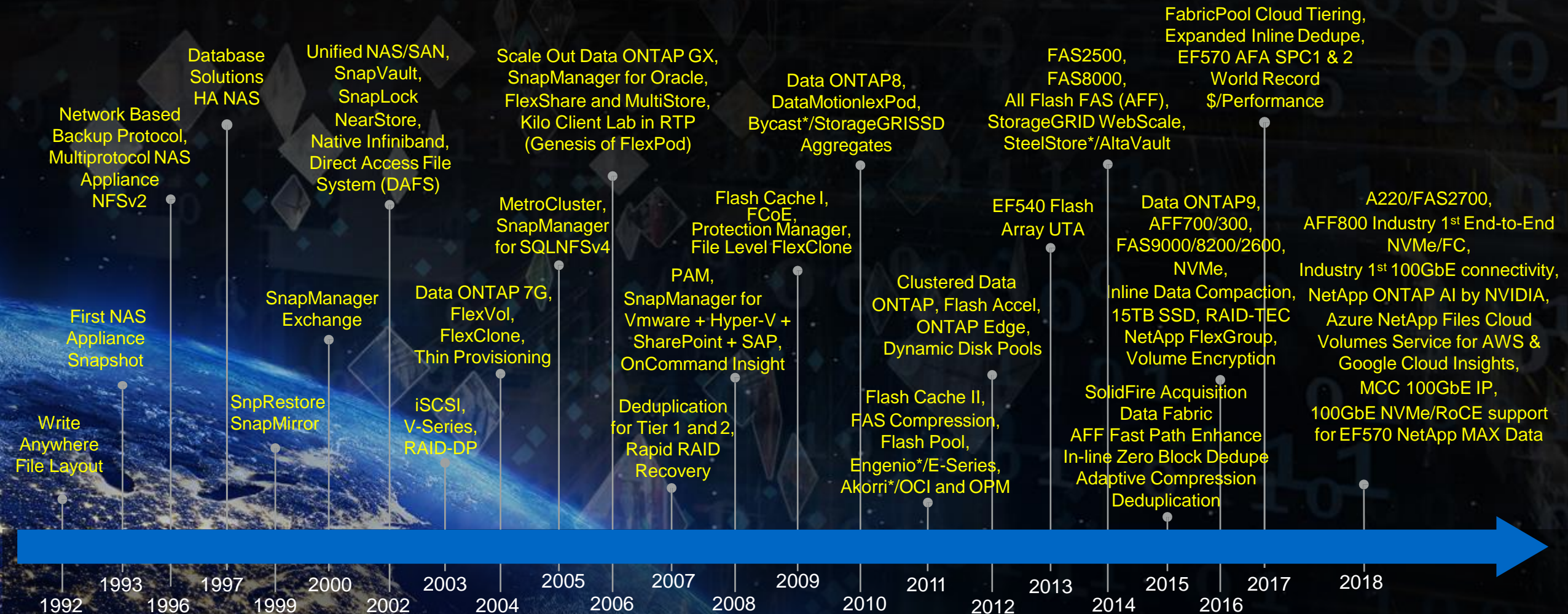


NetApp и трансформация хранения данных: новые носители и протоколы

Сюртуков Игорь
NetApp



NetApp: 26 лет инноваций



Magic Quadrant for General-Purpose Disk Arrays // 08.11.2018

<https://www.gartner.com/doc/reprints?id=1-3L2DG7Q&ct=161101>



Magic Quadrant for General-Purpose Disk Arrays Published 8 November 2018 - ID G00346291

Market Definition/Description

General-purpose disk arrays (GPDAs) satisfy the storage needs of applications and middleware, such as databases, backup/restore systems and archiving solutions running on physical or virtual servers. Block-and-file protocols — e.g., Fibre Channel (FC), Internet Small Computer System Interface (iSCSI), Network File System (NFS), Serial Attached SCSI (SAS) and SMB — dominate this market. NVMe-oF, an emerging protocol designed to enhance the performance of solid-state arrays (SSAs), will not be supported by hybrid disk arrays or all hard-disk drive (HDD) arrays, perhaps other than those with flash-centric architectures delivering submillisecond response times.





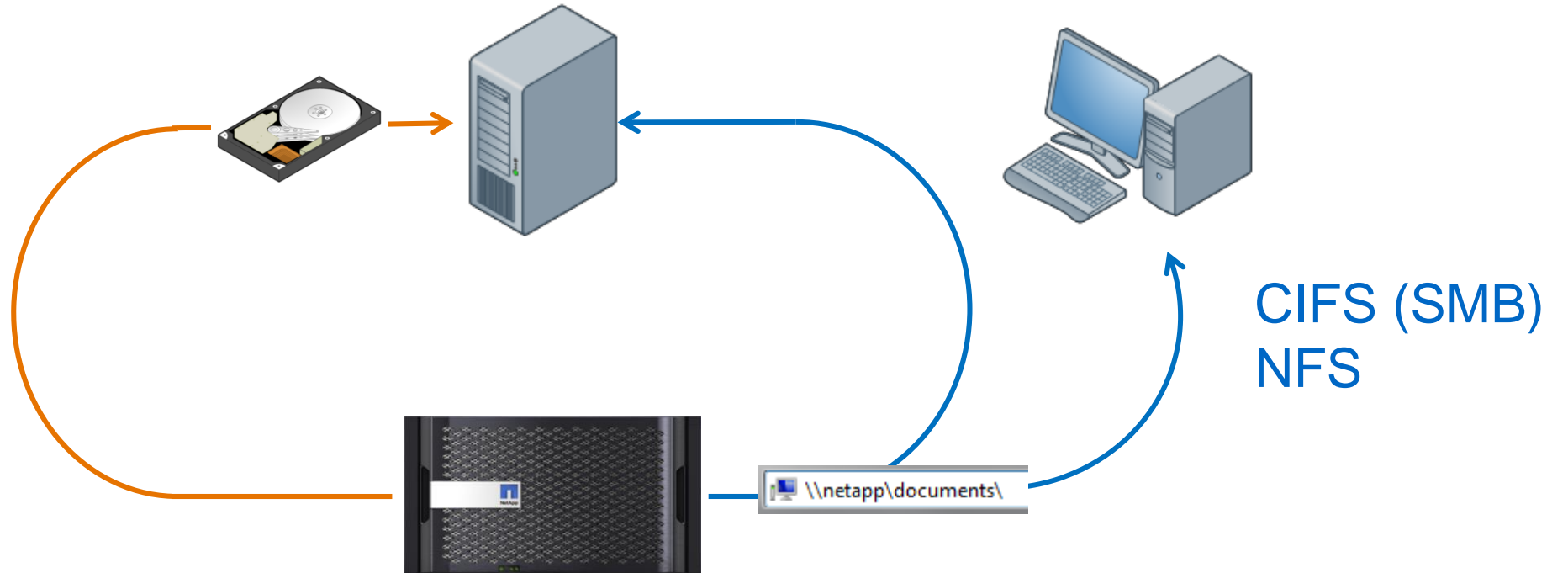
NetApp FAS/AFF и Cloud

Наш паровоз 

Унифицированный (unified) доступ

FC
FCoE
iSCSI

NVMe/FC
NVMe/RoCE
NVMe/InfiniBand



SAN

блочный доступ

NAS

файловый доступ



Data ONTAP was developed in 1992 by David Hitz, James Lau and Michael Malcolm.

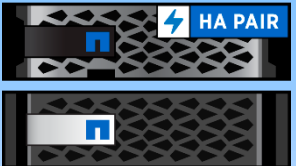
NetApp FAS/AFF – гибридные и all-flash СХД

A220
FAS2700




2 TB NVMe SSD
read cache onboard

A300
FAS8200



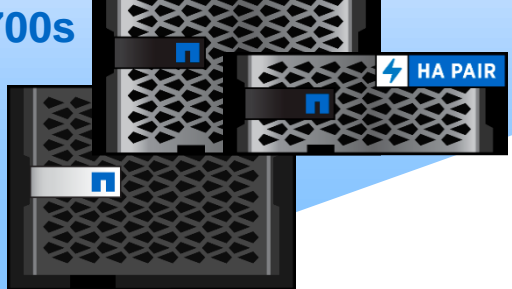
2-4 TB NVMe SSD
read cache onboard

A320




48 x NVMe

A700/A700s
FAS9000



2-16 TB NVMe SSD
read cache onboard

A800



48 x NVMe
+ SSD



Возможности NetApp ONTAP – базовые характеристики

- ❑ Поддержка защиты данных **RAID-DP** и **RAID-TEC** – любой классический RAID является компромиссом между надежностью хранения, производительностью и эффективности (объему дисков с данными в по сравнению с общим числом дисков в RAID), при этом у RAID-DP/TEC все эти показатели являются наилучшими:
 - **НАДЕЖНОСТЬ: RAID-TEC** – защищает от выхода из строя ТРЕХ ЛЮБЫХ дисков (уникально для рынка!), **RAID-DP** – от выхода из строя двух любых дисков. Для сравнения – RAID-10 гарантирует защиту от выхода из строя лишь одного диска,
 - **ЭФФЕКТИВНОСТЬ: RAID-DP** = 92,9% (диски SAS/SSD в RAID размером 28 дисков), **RAID-TEC** = 89,7% (29 дисков), для сравнения – эффективность RAID-10 = 50%,
 - **ПРОИЗВОДИТЕЛЬНОСТЬ: RAID-DP/TEC** – уровня RAID-10 или выше (что подтверждено промышленными тестами: <https://spcresults.org> и <http://www.spec.org>),
- ❑ Полноценная поддержка **всех современных протоколов доступа** (как блочных, так и файловых) – контроллеры NetApp унифицированы и на базе единой операционной системы **ONTAP** (ОС разработанная NetApp и развивающаяся уже более двадцати лет) и единого контроллера и осуществляется поддержка всех протоколов. Важно, что тут не используются РАЗНЫЕ модули (контроллеры) для разных протоколов, как у всех основных конкурентов. Поддерживаются протоколы NFS 3, 4 и 4.1, pNFS, SMB 2, 2.1 и SMB 3.1, iSCSI, FCP/FCoE и самый современный протокол NVMe (например, NVMe-over-FC даже с SAS SSD (не говоря уже о NVMe-носителях) даёт прирост производительности по сравнению с FCP в 50% и снижение задержек на 30%);
- ❑ **Кластеризация СХД** – возможность объединять до 12-и (SAN+NAS) или 24-х контроллеров (NAS) в рамках единой СХД – при этом контроллеры не должны быть одной модели или даже одного поколения (к примеру, развернув двухконтроллерную СХД сегодня возможно модернизировать ее добавлением контроллеров следующих поколений, которые будут выпущены на рынок NetApp в будущем);
- ❑ **Storage Virtual Machines (SVM)** – возможность разделения единой СХД на виртуальные системы хранения, каждая из которых имеет свои ресурсы (агрегаты, тома, виртуальные порты) и может администрироваться отдельно от всех других СХД (для повышения безопасности, для более гибкого управления ресурсами с помощью QoS и пр.);
- ❑ **Snapshot** – мгновенные снимки по технологии Redirect-on-Write (RoW), до 1023 (!) снимков на том без снижения производительности (не используется механизм Copy-on-Write) – это важно, т.к. многие конкуренты используют механизм CoW дающий многократное снижение производительности при использовании снапшотов;
- ❑ **Deduplication** – дедупликация для любых данных и типов дисков без влияния на производительность, включая дедупликацию датастора VMWare и блочных данных (томов, доступных через FC или iSCSI). Дедупликация не имеет скрытых ограничений и особенно привлекательна для виртуальных сред. Доступно два режима работы технологии дедупликации (inline и post-process);
- ❑ **Compression** – компрессия (сжатие) для любых данных и типов дисков, возможно как inline-сжатие, так и post-process-обработка. Доступно четыре режима работы технологии компрессии, в том числе специально разработанный для AFF-систем адаптивный режим компрессии, который для некоторых типов нагрузки не только уменьшает объем данных, а еще и увеличивает производительность,
- ❑ **Compaction** – упаковка для любых данных и типов дисков, позволяющая разместить сжатые и/или дедуплицированные данные таким образом, чтобы занимаемое место на дисках было минимальным. Процесс отработывает после дедупликации и/или сжатия и является еще одной уникальной ступенью NetApp в сторону уменьшения занимаемого пр-ва на СХД (появилось в ONTAP 9);
- ❑ **QoS (quality of service)** – управление производительностью для томов, LUN, файлов или целиком SVM – установка ограничений QoS Max. по IOPS или bandwidth или по IOPS/bandwidth одновременно, а также установка QoS Minimum (AFF и ONTAP Select Premium с SSD);
- ❑ **FlexGroup** – файловые контейнеры размером до 20 ПБ и до 400 миллиардов файлов с автоматической балансировкой нагрузки между всеми ресурсами кластера с высокой производительностью и отказоустойчивостью;
- ❑ **In-place upgrade** – возможность модернизировать СХД без переноса данных (NDU) – меняются только контроллеры, дисковые полки и данные на них остаются не тронутыми (без необходимости бакапа/восстановления). Систему можно модернизировать путем замены всех компонент (100%) на новое «железо» без прерывания доступа к данным;
- ❑ **Active IQ Unified Manager** – ПО мониторинга доступности, защиты данных и анализа производительности (более 200 счетчиков, построения графиков нагрузки с любым уровнем гранулярности);
- ❑ ...плюс множество дополнительных возможностей: BranchCache, VAAI/ODX, NDSR (nondisruptive shelf removal), FPolicy, Off-box antivirus и пр.

Возможности NetApp ONTAP – ещё немного...

...ещё немного технической информации:

- Операционная система ONTAP может работать не только на «железе», но и в виде Software Defined Storage (ONTAP Select) или облачного хранилища ONTAP Cloud Volumes в AWS, Azure, GCP (подробней: <https://cloud.netapp.com/ontap-cloud>);
- Поддерживается VAAI/VASA и интеграция в vCenter, VVols, а также SRM-адаптер для Vmware;
- Поддерживается ODX и интеграции в SCOM для Hyper-V, а также поддержка командлетов Windows PowerShell;
- Поддерживаются самые современные протоколы доступа, включая SMB 3.1.1 и NFS 4.1, а также протоколы авторизации и аутентификации;
- Интеграция в AD Windows (для файловых протоколов) с поддержкой всех прав и квот AD-пользователей + расширенная интеграция с помощью API FPolicy;
- Поддержка множества Nondisruptive Operations (NDO) – выполнение внутренних операций без прерывания ввода-вывода (LIF migrate, LUN/Volume move, Aggregate relocate, Storage failover);
- Поддержка Nondisruptive Upgrade (NDU) – выполнение апгрейда без прерывания ввода-вывода (добавление или замена контроллеров и полок, upgrade service processor/RLM/BMC, firmware update);
- Поддержка NDSR (nondisruptive shelf removal);
- Active IQ (AutoSupport) — возможность для СХД самостоятельно сообщать о проблемах с оборудованием в техподдержку NetApp, рекомендации по обновлениям ONTAP, прошивок дисков, дисковых полок, мониторинг производительности, рекомендации по оптимизции производительности и эффективности хранения.

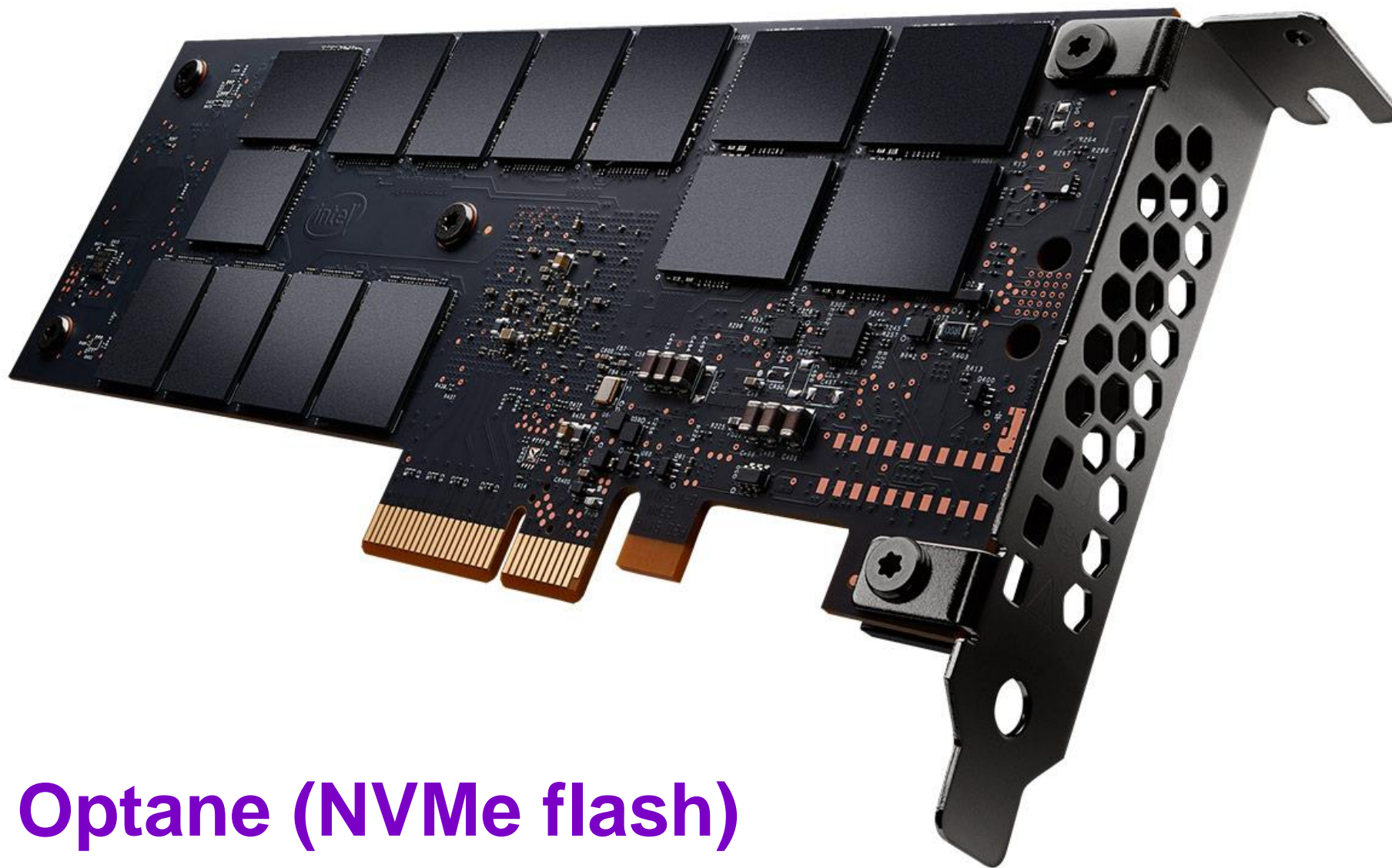
Возможности NetApp ONTAP – дополнительно

ДОПОЛНИТЕЛЬНО доступны следующие лицензии, не входящие в базовую поставку. Их можно приобретать в составе полного набора ПО под названием [Premium Bundle](#) (для AFF может называться Flash Bundle). Лицензии покупаются единожды и навсегда:

- ❑ [SnapRestore](#) – мгновенное восстановление (rollback) моментальных снимков SnapShot (если требуется функционал «откатывания» текущего тома на какой-то из снапшотов), сам же функционал моментальных снимков входит в базовую поставку и доступен без этой лицензии в полном объеме,
- ❑ [SnapMirror](#) – удаленная **асинхронная** и **синхронная** репликация на другую систему хранения FAS – решение DR (Disaster Recovery), с сохранением эффективности дедупликации и компрессии;
- ❑ [SnapVault](#) – удаленная репликация снапшотов на любую другую систему хранения FAS для целей disk-to-disk backup, с сохранением эффективности дедупликации и компрессии,
- ❑ [FlexClone](#) – гибкие клоны (снапшоты с возможностью записи),
- ❑ [SnapCenter](#) – централизованное создание консистентных снапшотов и клонов для тестовых сред, а также для резервного копирования. Доступные плагины:
 - SnapCenter Plug-In for Microsoft Windows, SQL Server, Exchange Server,
 - SnapCenter Plug-in for SAP HANA Database,
 - SnapCenter Plug-In for VMware vSphere,
 - SnapCenter Plug-In for Oracle Database,
 - SnapCenter Plug-in for MongoDB (community support),
 - SnapCenter Plug-in for DB2 (community support),
 - SnapCenter Plug-in for MySQL (community support),
 - SnapCenter Plug-in for SAP MaxDB (community support),
 - SnapCenter Plug-in for SAP Adaptive Server Enterprise /SAP ASE/ (community support),
 - ...дополнительно поддерживается бесплатный фреймворк для написания плагинов интеграции снапшотов с любыми приложениями. Для написания плагинов можно использовать Perl, PowerShell, Java, UNIX Shell и Python.
- ❑ [Single Mailbox Recovery](#) – восстановление отдельных почтовых ящиков или сообщений в среде Microsoft Exchange,

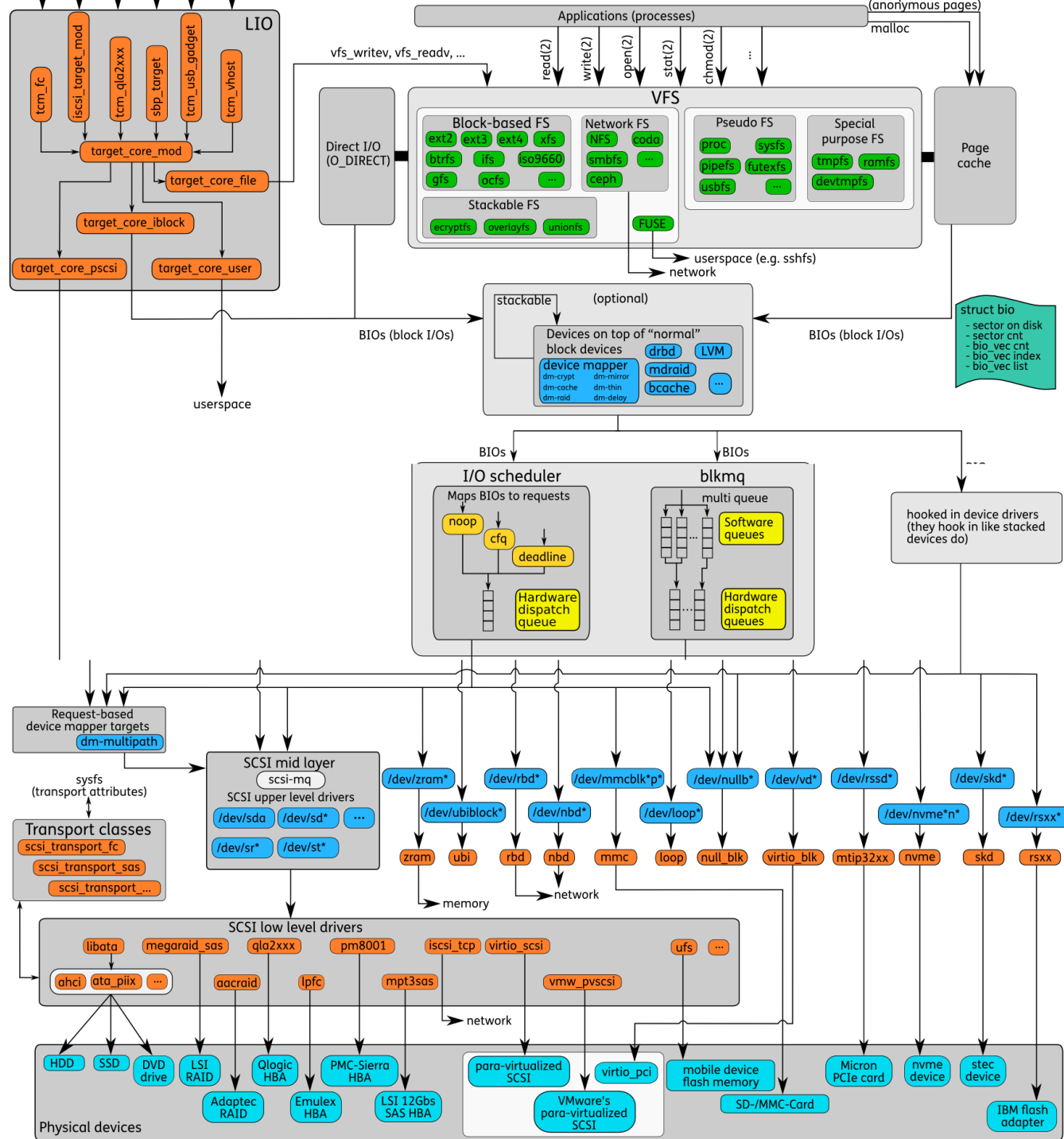
- ❑ Также доступны дополнительные лицензии, **не входящие** в базовую поставку или Premium Bundle:
- ❑ [SnapLock](#) – функционал блокировки файлов Write-once/Read-many (WORM) для CIFS/NFS – гарантированное хранение в рамках установленного срока, без возможности удаления (соотв. стандартам SEC 17a-4(f) rule, FINRA, CFTC),
- ❑ [FlexArray](#) – виртуализация СХД сторонних производителей (не поддерживается младшими моделями NetApp). Сторонние СХД подключаются к NetApp FAS и используются, как собственная дисковая емкость со всеми преимуществами ONTAP,
- ❑ [FabricPool](#) – оптимизация расходов на хранение данных на SSD при помощи автоматического тиринга (переноса) "холодных" данных в объектное хранилище по протоколу S3. Работает как внутри дата-центра, так и в публичные облака (AWS, MS Azure, GCP, IBM Cloud, Alibaba),
- ❑ [FlexCache](#) – кэширование файловых данных (NFSv3) для удаленных СХД, используется для ускорения чтения больших объемов данных, снижения задержек при доступе через WAN. В качестве кэширующих СХД могут выступать любые модели, работающие под управлением ONTAP, в том числе ONTAP Select и NetApp Cloud Volumes ONTAP,
- ❑ [MAX Data](#) (Memory Accelerated Data) – обеспечение сверхнизких задержек за счёт тиринга (переноса) данных между AFF-системами и модулями Intel Optane DC PMM в серверах. Обеспечиваются задержки в районе 0,02 мс (20 микросекунд) на чтение и запись, с поддержкой функциональности создания снапшотов на системах AFF и репликации этих снапшотов на другую систему FAS/AFF.

Что это?

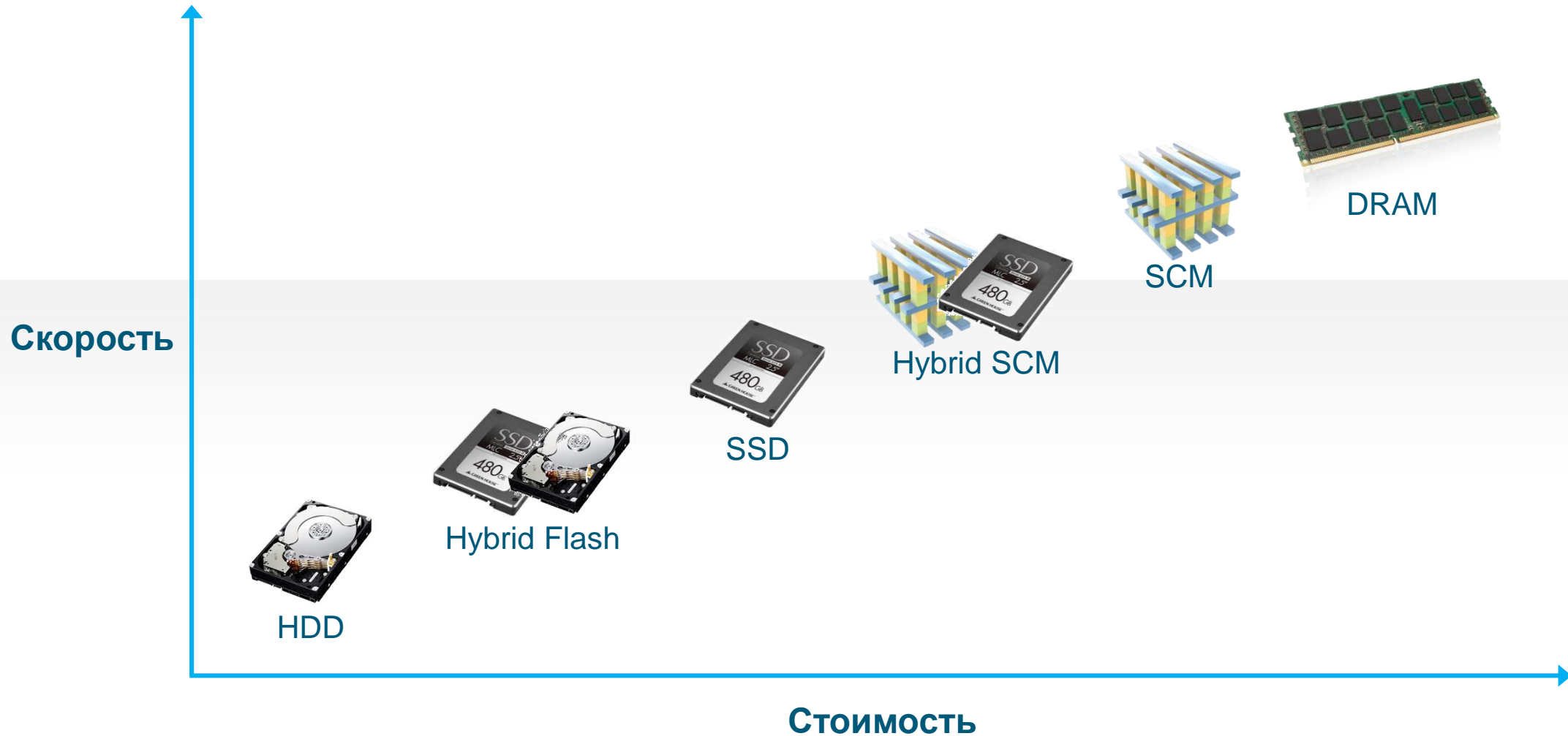


Intel Optane (NVMe flash)

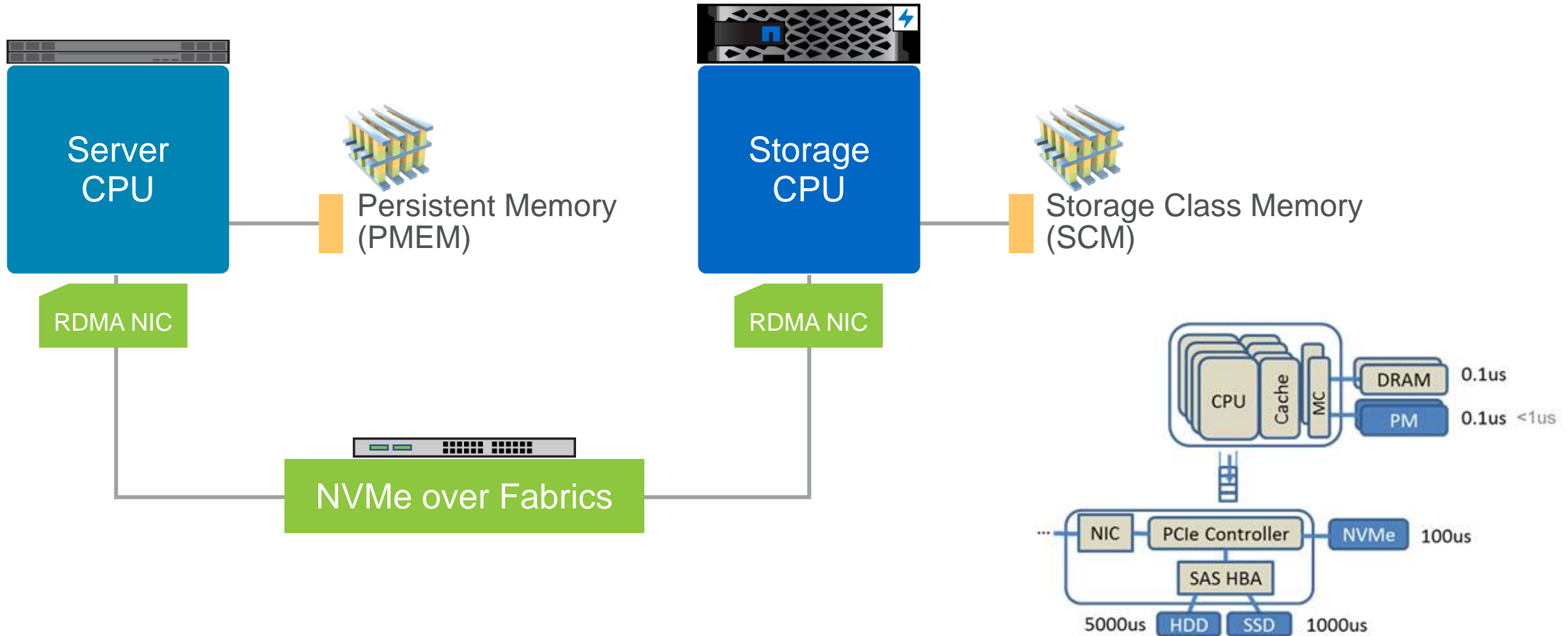
Что это?



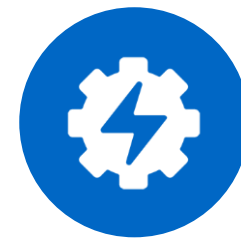
Прогресс носителей данных



Подход для NVMe-oF и Storage-Class Memory (SCM)



AFF A320 – end-to-end NVMe в сегменте midrange



Ускорьте самые требовательные приложения:

- Ультранизкая задержка: **~0,1 ms**
- **380 000 IOPS** с задержкой **0,3 ms** (Oracle SLOB)
- Пропускная способность:
 - **12 GB/s reads** для HA-пары
 - **5 GB/s writes** для HA-пары

2U / 24 x NVMe SSD

max: 48 x NVMe SSD

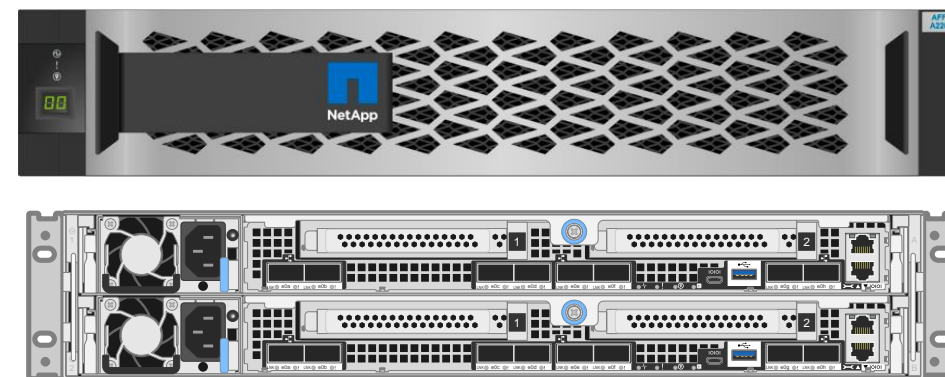
Масштабирование:

- До 48 NVMe-SSD (полки: 400Gb/s NVMe/RoCE v2)
- До 24-х узлов (12 HA-пар)

Широкий спектр рабочих нагрузок:

- Пропускная способность и масштабирование для deep learning
- IOPS и низкая задержка для machine learning
- Ядро высоконагруженных ЦОД и облачных сервисов

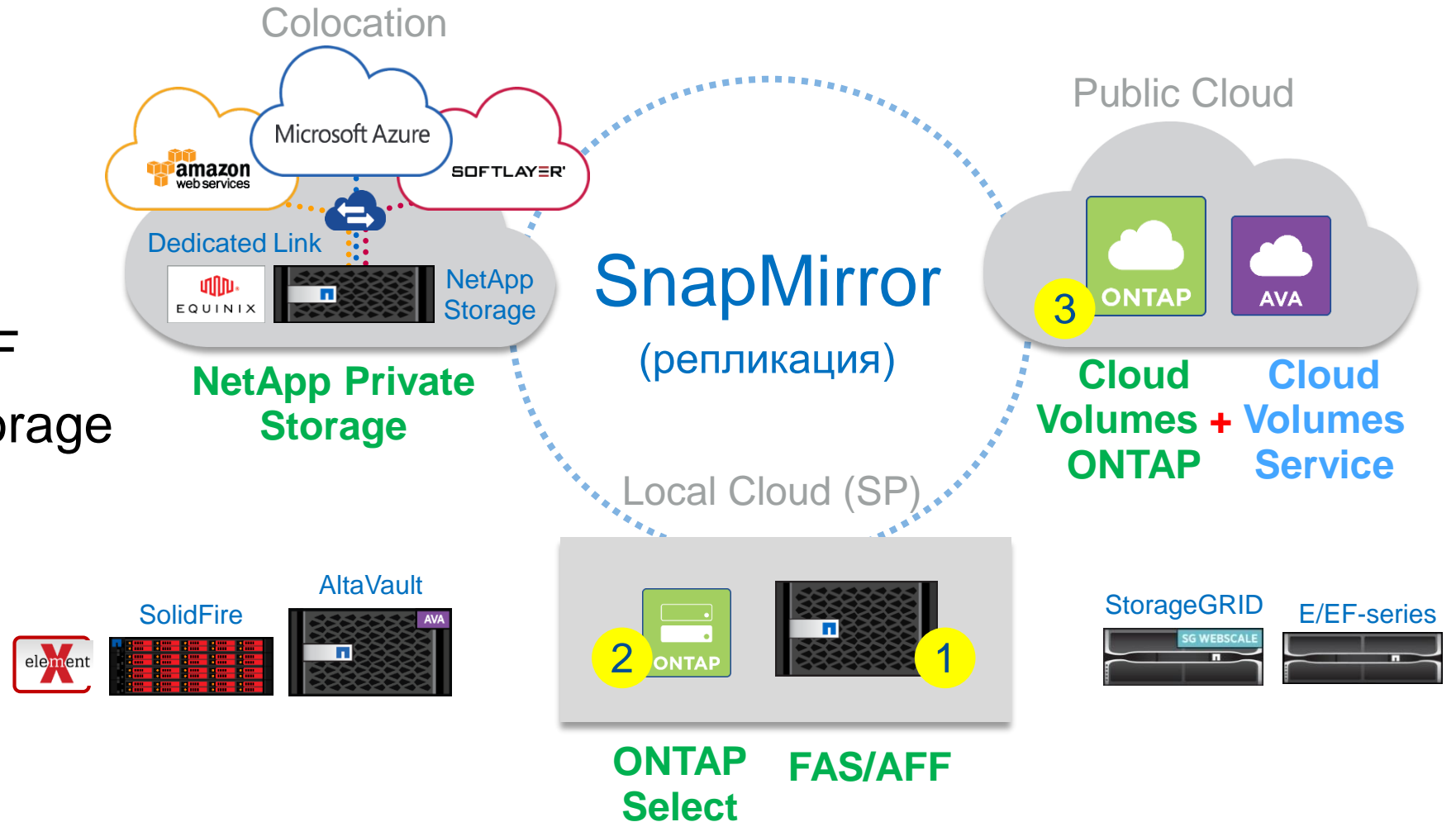
- 100GbE RoCE (100/40/25/10)
- 32Gb FC (32/16/8)



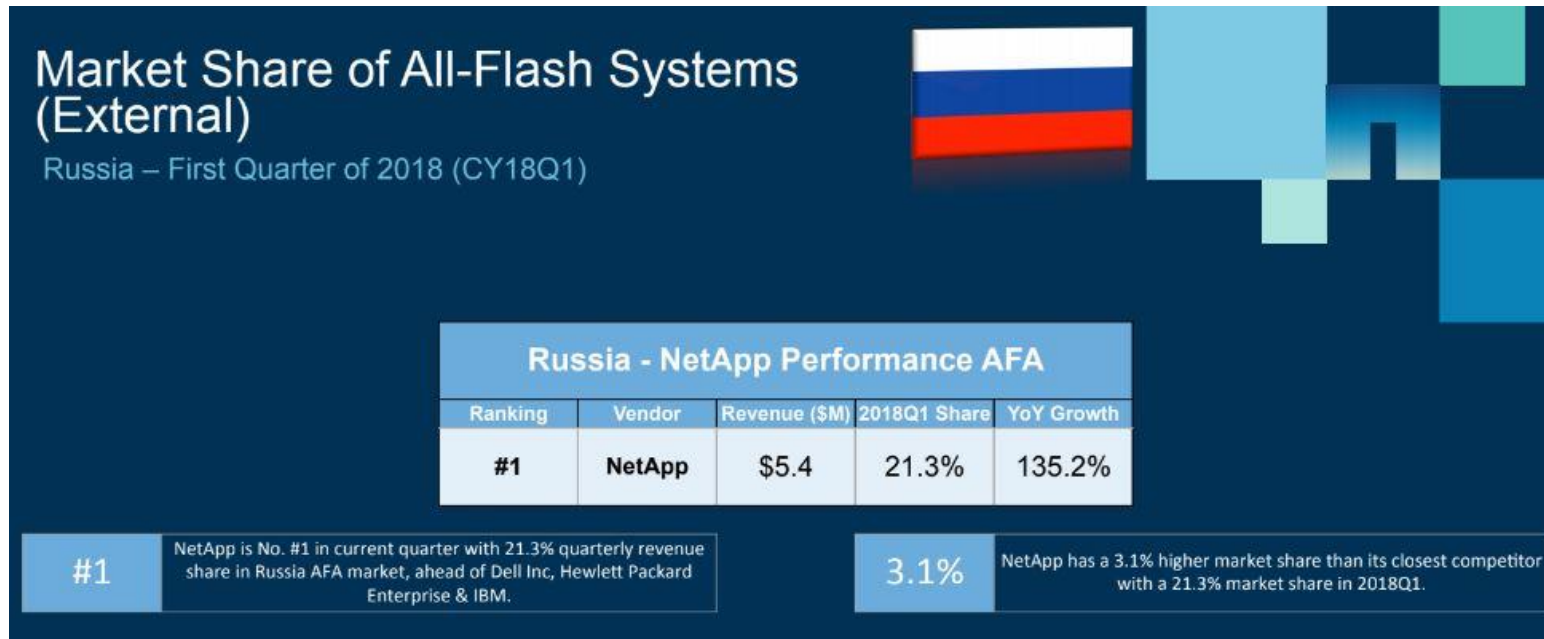
Модель NetApp Data Fabric

Единая ОС ONTAP:

1. СХД серии FAS/AFF
2. Software-defined storage
3. Cloud storage



Market Share of all-flash systems (external)





Серия NetApp EF (all-flash)

- Просто
- Недорого
- Очень быстро

NetApp EF600 (all-flash) – очень просто и недорого

Простая и производительная архитектура:

- Два контроллера:
 - ✓ 100Gb NVMe/InfiniBand
 - ✓ 100Gb NVMe/RoCE
 - ✓ 32Gb NVMe/FC + 32Gb FC
- До 24 шт. NVMe-SSD (от 1,9 до 15,3 TB)
- > 2 000 000 IOPS (4K read 100%)
- > 42 GBps (гигабайт в секунду, read 100%)
- Задержки < 0,1ms (для определенных нагрузок)
- Всегда полный пакет ПО (бесплатно)
- Доступность 99,9999%



NetApp EF600

Анонс: Q3Y2019

Тестирование модели **предыдущего поколения EF570** (Q3Y2017) с рекордными результатами:

Тест **SPC-1** для **24 x 800GB SSD** – результат **500 022 IOPS** при задержке **0,44 ms**

- https://spcresults.org/sites/default/files/files/executive_summary/SPC-1_A31009_ES_NetApp-EF70%20v2.pdf

Тест **SPC-2** для **20 x 800GB SSD** – результат **20,6 GBps** read / **6,4 GBps** write

- https://spcresults.org/sites/default/files/files/executive_summary/B12003_ES.pdf

The END

Неофициальный канал NetApp



<https://t.me/storagetalks>