

## Распределённое масштабируемое решение VDI MxGPU

### Состав

- Серверы MxGPU (VDI) рекомендуемое количество – не менее 2
- Серверы хранения образов (ZFS) рекомендуемое количество – 2
- Внутренняя сетевая среда (iNW)  
Ангара; 25/10 GbE RoCE; 1GbE управления
- Внешняя сетевая среда (eNW)  
2x GbE

### Базовая машина

Применяется для VDI, а для единообразия может применяться и для всех остальных серверов.

Может самостоятельно выполнять функциональные возможности хранения образов Виртуальных Машин (VM).

Сервер в 2U исполнении с одним ЦПУ **Ерус 7551P**, **32 ядра** и 8 каналами оперативной памяти (со 2 квартала 2019 – 64 ядра с поддержкой PCIe v4).

### Состав базового решения:

- **256ГБ** оперативной памяти DDR4, 2600 Reg, ECC;
- **2x GPU AMD Radeon™ Pro V340**  
или AMD Radeon™ Pro **7150x2**
- **6x** шпиндельных дисков SATA **10/12ТБ**  
с возможностью замены в горячем режиме
- **4x SSD** дисков **NVMe 1ТБ**;
- **2x SSD** диска SATA 120ГБ с горячей заменой;
- **1x** сетевой интерфейс Ангара с поддержкой RDMA;
- **2x** порта 25GbE RoCE с поддержкой RDMA;
- **3x** порта 1 GbE (один с IPMI);
- **2x** порта USB 3.0, **2x** USB 2.0, **1x** COM, **1x** VGA;
- **2x** блока питания (1+1) 770Вт с возможностью горячей замены;
- **8x** вентиляторов охлаждения бсм с заменой в горячем режиме.



### Оptionальные возможности:

- Поддержка до 2ТБ ОЗУ
- 128x lane PCIe G3/4\* с размещением в 6 стандартных слотах 8x lane или в 5 стандартных слотах +1 мезонин (8x) - 2 слота из общего числа стандартных могут поддерживать до 16x lane
- до 32x портов SATA (до 15 устройств в каждом) или NVMe

### Машина 1U

Для серверов VDI, которым требуются большие мощности GPU можно применять сервера в 1U исполнении с 4x посадочными местами под GPU с x 16x lane PCIe слотами и 2x БП (1+1) 1600Вт.

### Рекомендуемый состав:

- **512ГБ** оперативной памяти DDR4, 2600 Reg, ECC;
- **4x GPU AMD Radeon™ Pro V340**  
или AMD Radeon™ Pro **7150x2**
- ЦПУ **Ерус 7551P**, **32 ядра** (со 2 квартала 2019 – 64 ядра с поддержкой PCIe v4);
- **2x SSD** диска **NVMe 1ТБ**;
- **2x SSD** диска SATA 2ТБ с горячей заменой;
- **2x SSD** диска SATA 120ГБ с горячей заменой;
- **1x** сетевой интерфейс Ангара с поддержкой RDMA;
- **2x** порта 25GbE RoCE с поддержкой RDMA;
- **3x** порта 1 GbE (один с IPMI);
- **2x** порта USB 3.0, **2x** USB 2.0, **1x** COM, **1x** VGA.

## Основные функциональные возможности системы VDI:

Решения на основе

- VMware Horizon 7.
- Citrix XenDesktop 7.15+.
- KVM.

Принципиально новым в решении является применение **NVMe устройств** и возможностей **RDMA**. Основная цель NVMe – снятие имевшихся ограничений на очереди: если предшественник, AHCI, поддерживал только 1 очередь CHS на устройства с глубиной в 32 команды, NVMe поддерживает 64k очередей с глубиной 64k команд каждая. При этом NVMe реализует прямой доступ к данным, избегая наследования иерархии геометрии шпиндельных дисков. Набор команд сокращён до 13, упрощая и унифицируя технологию обмена. Все команды могут напрямую выставляться в PCIe или иную среду обмена (например, Ангара, RoCE). В качестве справки, приведём порядковые значения скоростей последовательного чтения/записи в потоке:

- 0.25ГБ/с -шпиндельный диск
- 0.5ГБ/с -SSD SAS/SATA
- >3.5ГБ/с -NVMe SSD

Внутренние операции поддержки СХД обсуждаемого масштаба задач предоставления образов VDI в согласованном состоянии требуют полосы пропускания между узлами, *превышающей* возможности 10GbE.

Основным преимуществом аппаратной виртуализации является то, что оборудование управляет циклами расписания, доставляя предсказуемое качество обслуживания (QoS). Такие фиксированные циклы расписания делятся в равной степени делятся на все VF, гарантируя то, что каждая VF получит свою справедливую долю служб GPU.

Предсказуемые производительность или детерминированное QoS приводят к плавному переходу от пилотных проверок концепции к оснащениям в масштабах всей организации. Диспетчеры пилотов определяют имеющиеся возможности GPU во время стадии проверки концепции и масштабируют вверх или вниз плотность пользователей (число пользователей на GPU) по мере необходимости.

Способность определять потребности GPU в базе данных пользователей зависит от возможностей организации прогнозировать и планировать свои ресурсы. При недостаточном прогнозировании результатов мы получаем в результате неспособность удовлетворения ожиданий пользователей; при чрезмерном прогнозировании результатов приводит к неполному использованию конфигурации.

Предсказуемый характер решения Multiuser GPU помогает избегать обоих этих нежелательных последствий.

### Дополнительные материалы:

<http://sddc.mdl.ru>

<http://angara.nicevt.ru/angaraK1/>

<http://onreader.mdl.ru/MxGPU/AMD.MxGPU.WhitePaper/content/index.html>

<http://www.mdl.ru/Solutions/Put.htm?Nme=MdlStorage#RX026a>