



Энергоэффективные и высокоплотные решения для ЦОДов

Юрий Мигаль

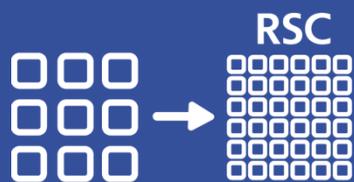
Руководитель департамента внедрения и эксплуатации

2023

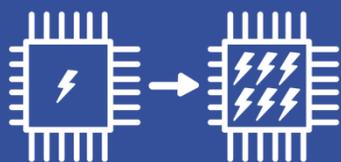
Высокопроизводительные системы с 2009 года

Разработка инновационных, энергоэффективных, высокопроизводительных и высокоплотных вычислительных систем для решения уникальных задач

Ключевые требования к НРС



Вычислительная
плотность



Энергетическая
плотность



Энергоэффективность



Легкость
управления и
обслуживания

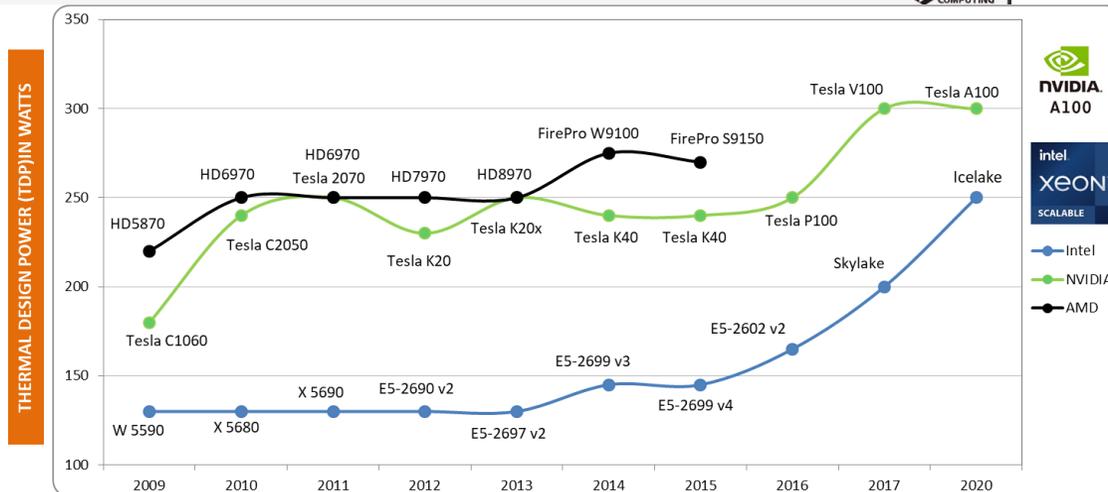
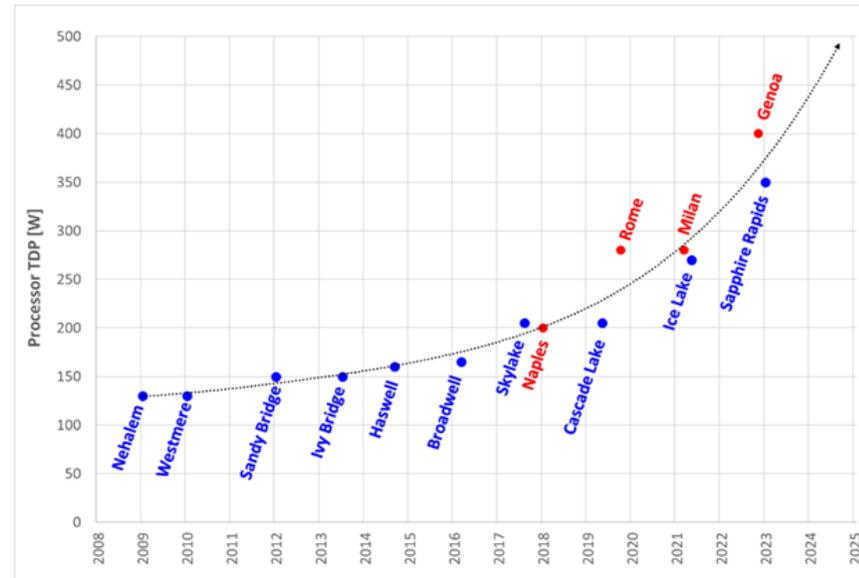
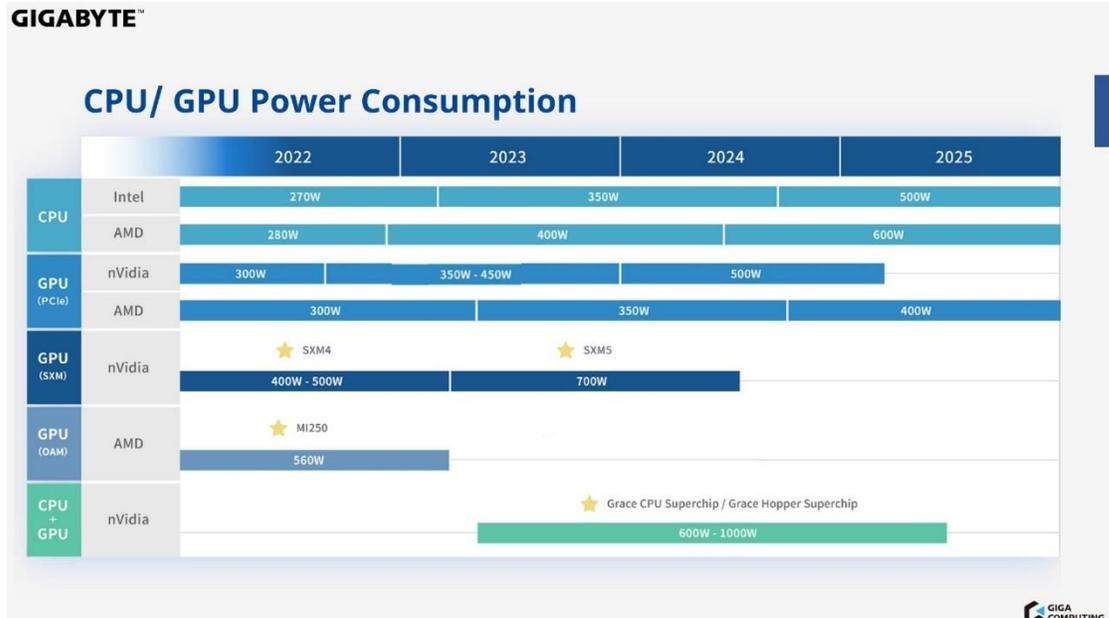


Надежность

Тенденции современных ЦОД

- Увеличение мощности единицы оборудования
- Увеличение нагрузки на стойку
- «Зеленость»
- Увеличение температуры ЦОД
- Снижение PUE

Рост потребления CPU и GPU не остановить



Liquid Cooling Technology

~50/50 Demand Split

Immersion

Single Phase



- PA06: Zero GWP, cheaper
- FC-40 & alternatives: Higher GWP, but better cooling
- Material Compatibility

Two Phase



- FC-3284: Better Cooling
- High GWP
- Material Compatibility
- Heatsink design is boiling enhancement coating

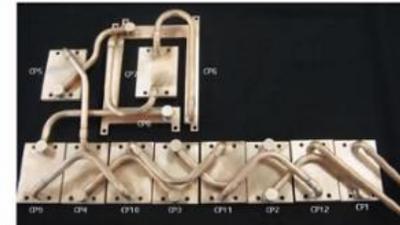
Cold Plate

Single Phase



- PG25: Better Cooling
- Zero GWP
- Parallel connections to avoid pre-heat

Two Phase



- R134a, Novec7000, or other refrigerant
- Better cooling
- Enables dense systems
- Series connections ok

More Niche

Сравнение погружного охлаждения

Драйверы
однофазного
охлаждения

	Однофазное	2-х фазное
Тип жидкости	Синтетическое углеводородное масло Фторуглерод (высокий ПГП)	Фторуглерод Фторкетон
Токсичность	Низкая токсичность	Проблема токсичности (Вытеснение кислорода)
ПГП (потенциал глобального потепления)	Низкий	Высокий
Обслуживание	Обслуживание систем в режиме реального времени	Автономное обслуживание
Стоимость	Низкая относительная стоимость	Высокая относительная стоимость
Герметичность	Отсутствие избыточного давления в ванне	Ванна под давлением
Состояние	Только жидкость	Жидкость и газ
Вязкость	Выше чем у воды	Ниже чем у воды

Жидкость придет и к Вам. Вопрос – КОГДА?

https://www.supermicro.com/en/solutions/liquid-cooling

Liquid Cooling: Efficient Thermal Management

- 2U2N BigTwin® SYS-221BT-DNTR > **1+1 x 2200Вт**
- 2U4N BigTwin® SYS-221BT-HNTR > **1+1 x 3000Вт**
- 8U 20N SuperBlade® SBI-421E-1T3N >
- 1U Hyper SYS-121H-TNR > **1+1 x 1200Вт**
- 2U Hyper SYS-221H-TNR >
- 2U Hyper AS-2125HS-TNR >
- 4U GPU SYS-421GU-TNXX >
- 4U PCIe GPU SYS-421GE-TNR >
- 4U PCIe GPU AS-4125GS-TNRT >
- 8U 8GPU SYS-821GE-TNHR >
- 8U 8GPU AS-8125GS-TNHR >
- 4U8N FatTwin® SYS-F511E2-RT > **2+2 x 2000Вт**
- 4U4N FatTwin® SYS-F521E3-RTB >

https://www.gigabyte.com/Industry-Solutions/coolit-liquid-cooled-ready-servers

Advantage | Applications | Products | FAQ | Partners | Resources | **Email Sales**

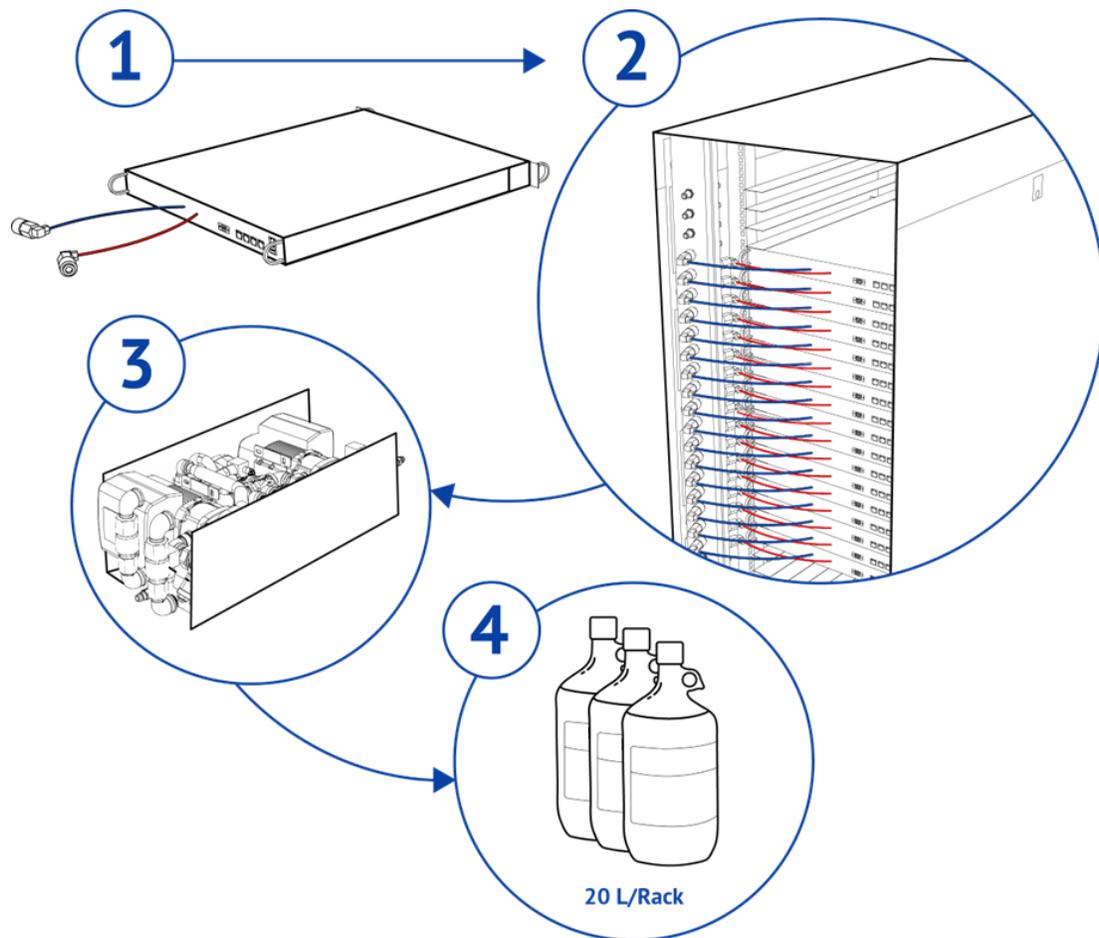
Related Products

- 3+1 x 3000Вт**
CoolIT DLC Solution
G492-ZL2
CPU/ SXM4 GPU
[Learn More](#)
- 1+1 x 3000Вт**
CoolIT DLC Solution
G262-ZL0
CPU/ HGX A100 GPU
[Learn More](#)
- 1+1 x 2200Вт**
CoolIT DLC Solution
H262-ZL0
CPU
[Learn More](#)
- 1+1 x 2200Вт**
CoolIT DLC Solution
H262-ZL2
CPU/ RAM/Networker
[Learn More](#)

ASUS SERVERS & WORKSTATIONS

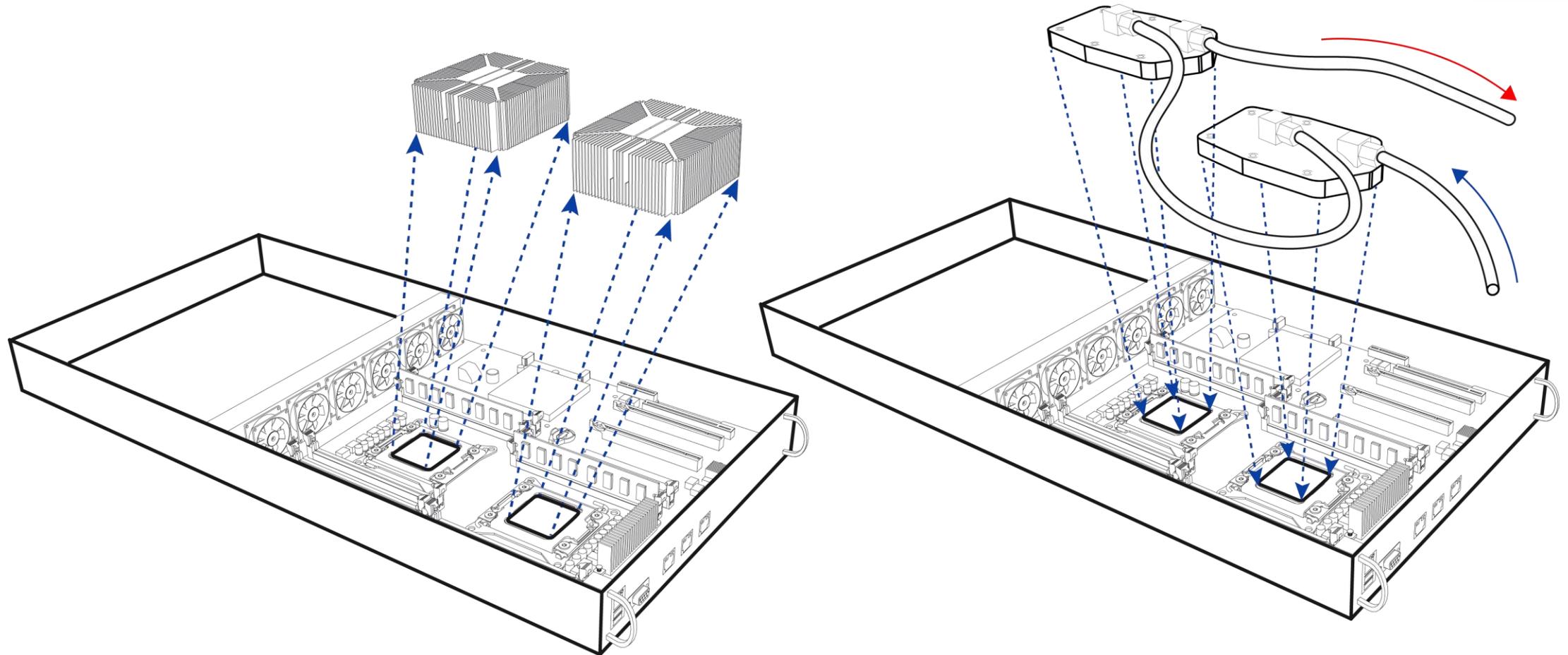
- 1+1 x 2600Вт**
- 1+1 x 3200Вт**

Решение РСК для ЦОД и облачных платформ – RSC HybridCoolingSolution (HCS)



- Работает с сервером **любого** производителя
- До **70%** уменьшения OPEX на охлаждение по сравнению с традиционным охлаждением воздухом
- Уменьшенное энергопотребление за счет снижения работы вентиляторов в серверах, отсутствия воздушных кондиционеров и фрикулингу (свободному охлаждению)
- Существенное сокращение использования площади ЦОД за счет более плотного размещения серверов в стойке (**до 4х раз, с 8-10кВт до 45кВт**)
- Возможность повысить температуру в серверной
- Возможность переиспользования тепловой энергии, отводимой от ЦОД

RSC HybridCoolingSolution: Установка набора для сервера



- Модули RSC HybridCoolingSolution заменяют стандартные радиаторы воздушного охлаждения и монтируются в те же отверстия на плате сервера

Узлы регулирования (насосные станции) для вычислительных систем

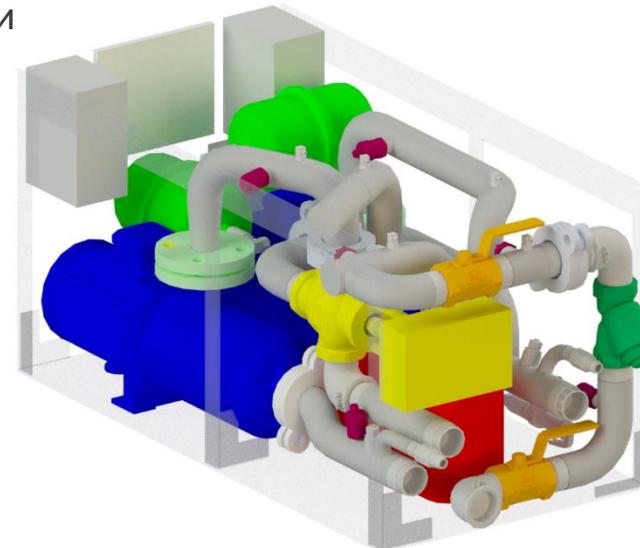
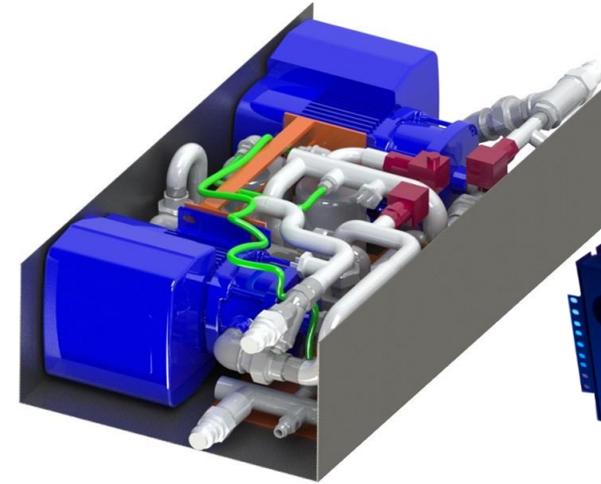
Более 10 лет опыта применения жидкостного охлаждения в ЦОД

Теплообменно-насосные станции обеспечивают перенос тепловой энергии и разделяют гидравлические контура с различными температурными, тепловыми, гидравлическими режимами или с различными типами жидкостей

Область применения

Работает с:

- Охлаждающие пластины РСК
- Охлаждающие пластины сторонних производителей



Отдельно стоящие узлы регулирования

Отдельно стоящие насосные станции РСК представляют собой независимые изделия, в форм-факторе стойки. Позволяют отвести тепловую мощность до 1000 МВт.



CDU 400



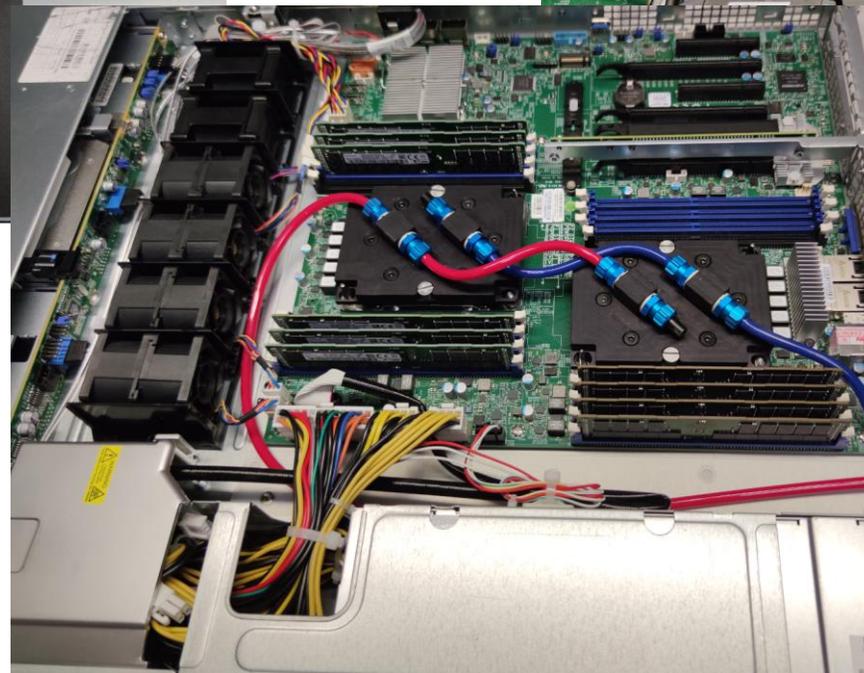
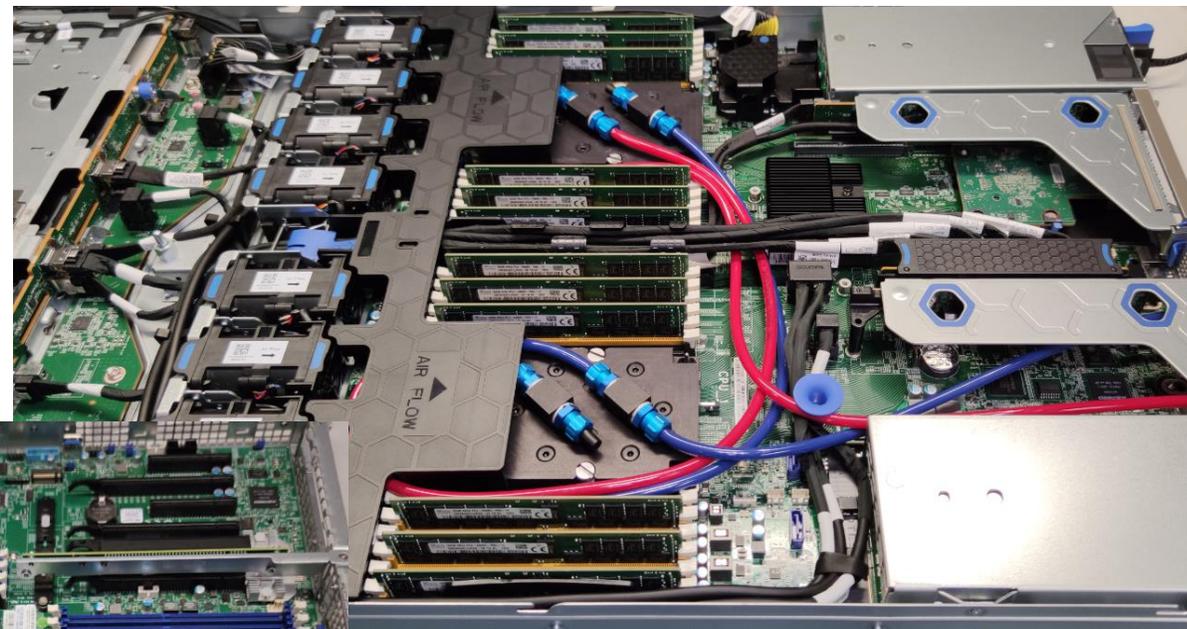
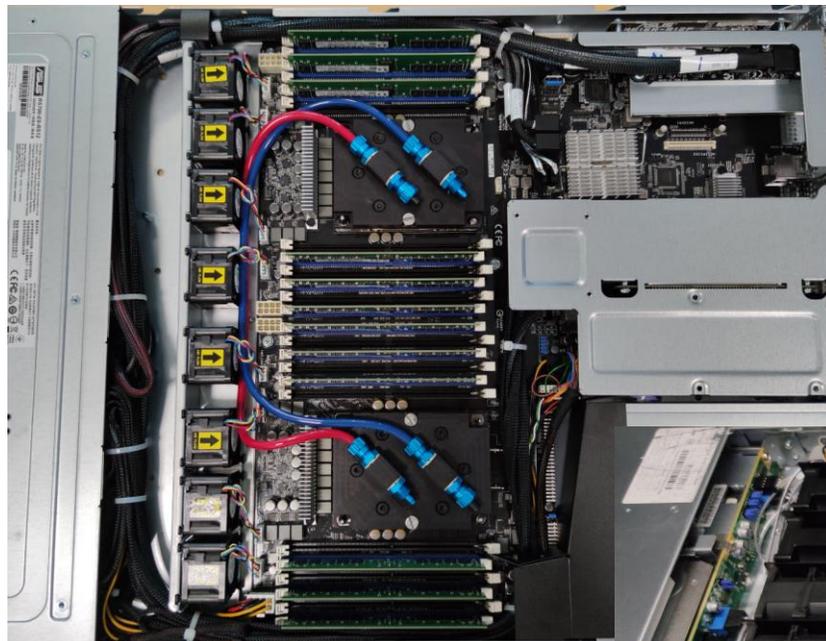
CDU-GM 400



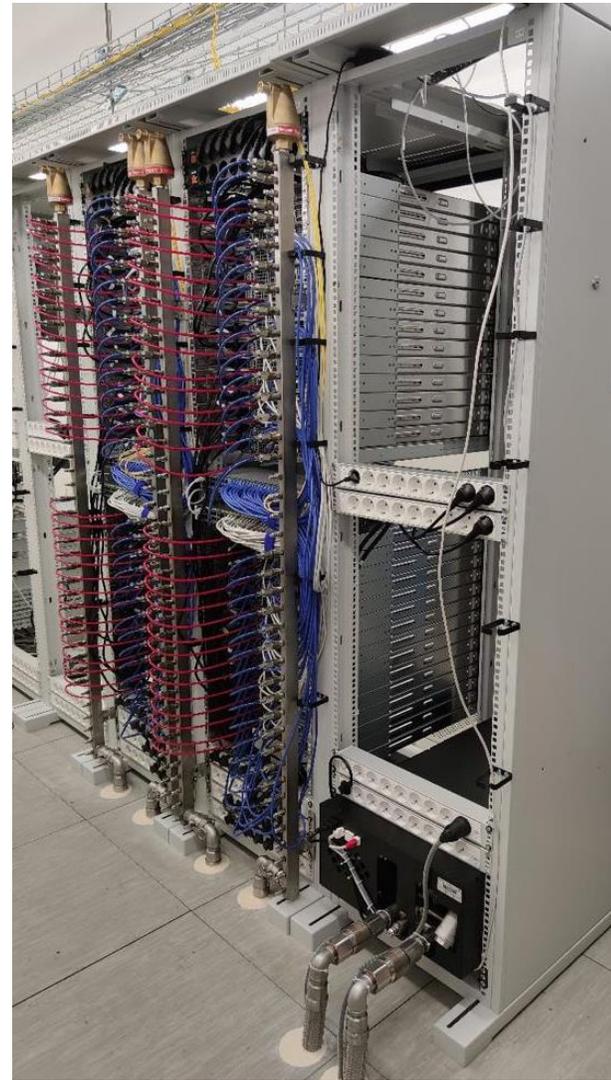
CDU-GM 600



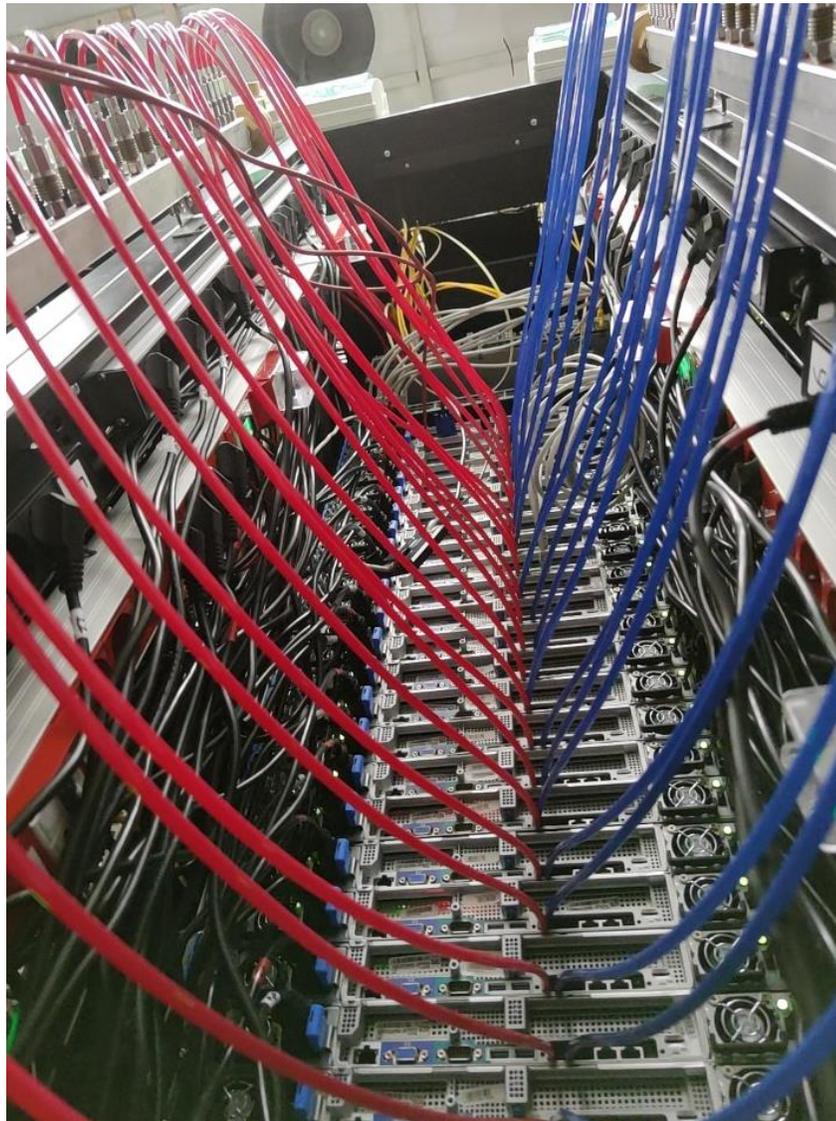
Пример инсталляции для Intel Xeon 2-го поколения



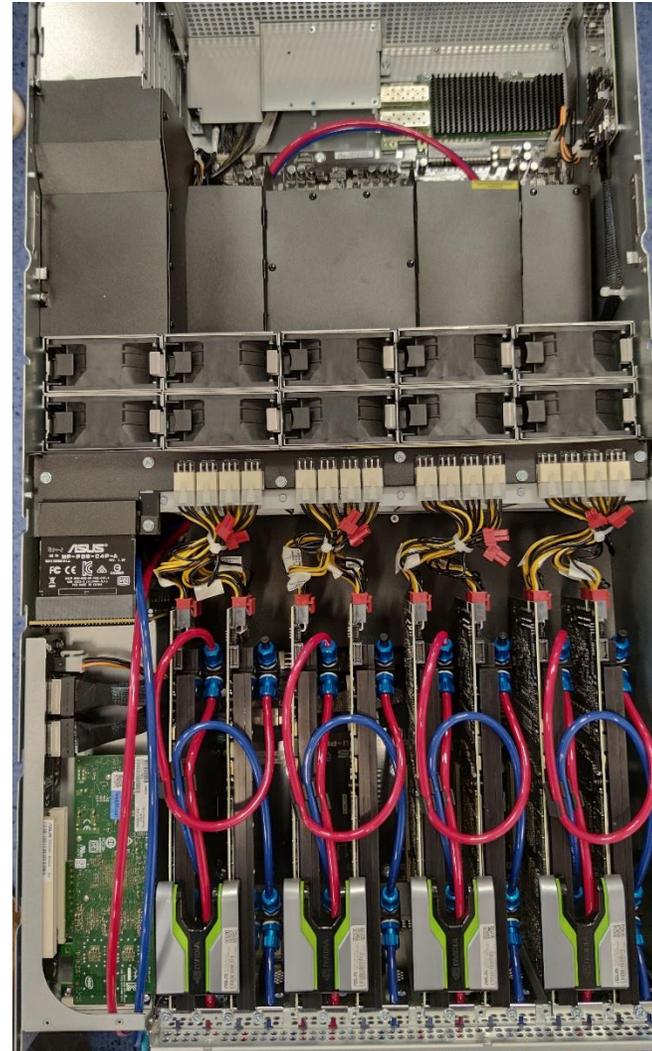
Первый этап проекта M100 VK (Москва, 2020)



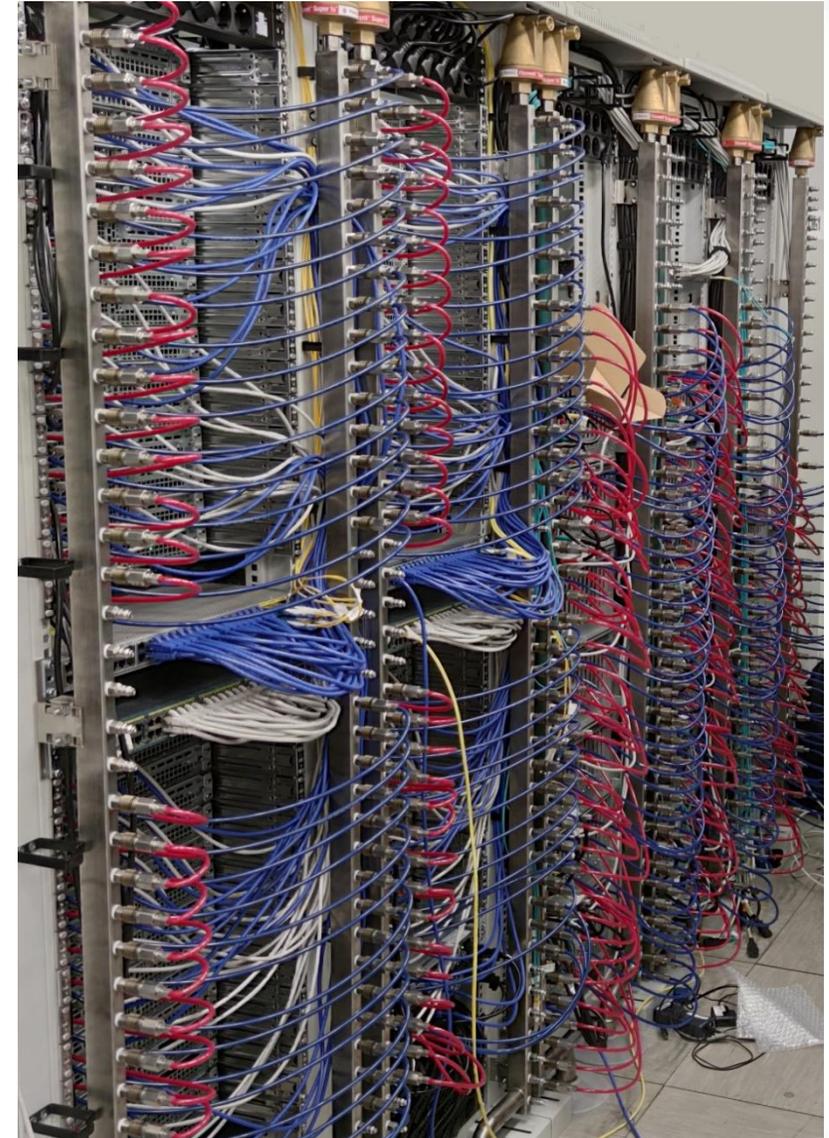
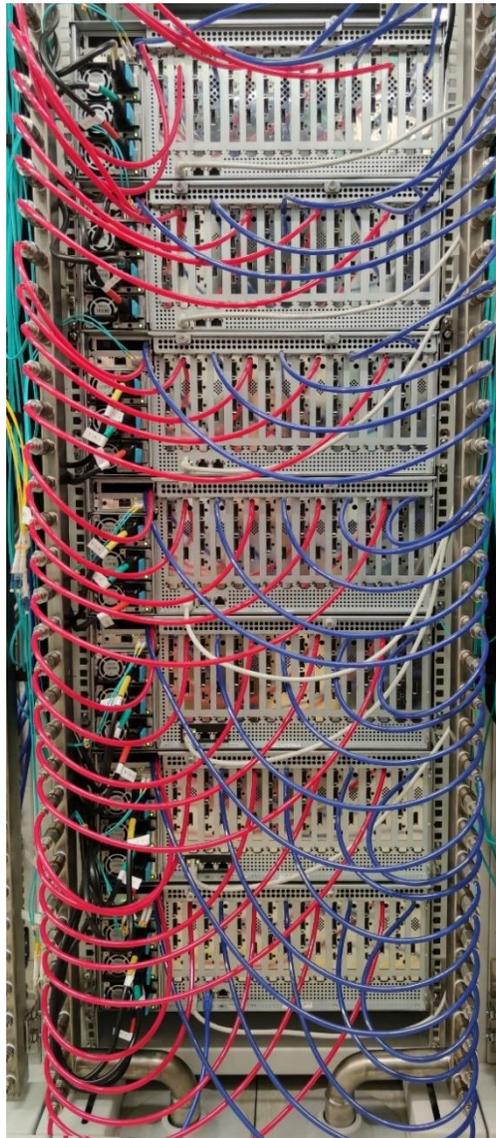
VK ИЦВА (2020)



Охлаждение ускорителей RTX8000



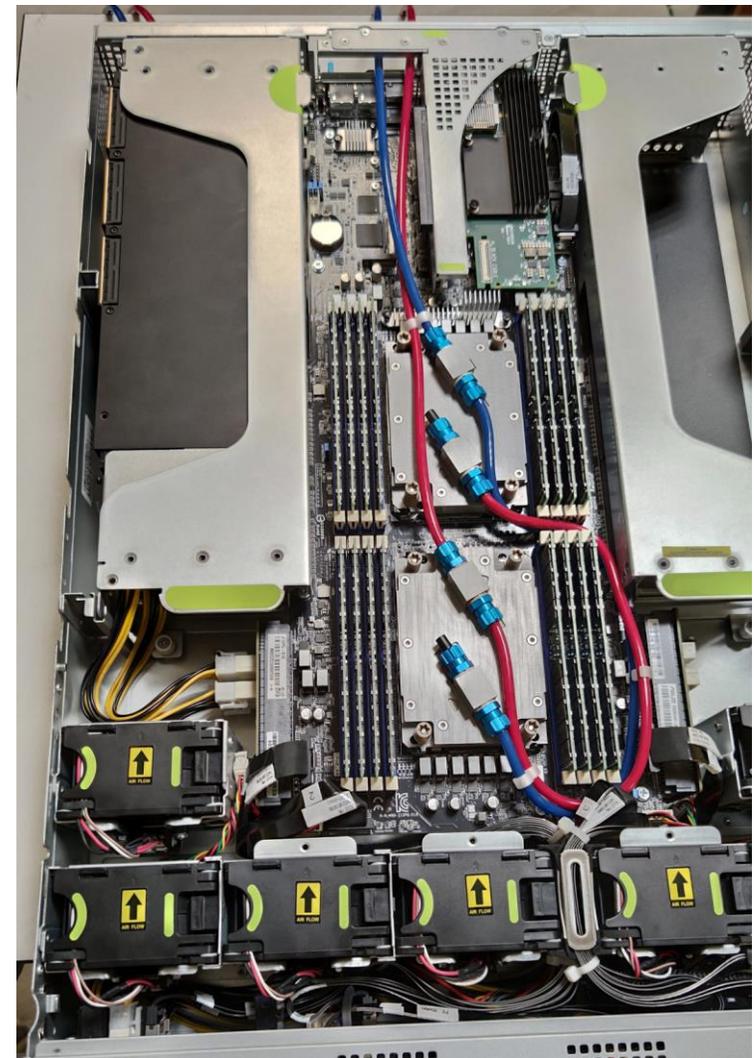
Второй этап проекта M100 VK (Москва, 07.2021)



Результаты внедрения

	Воздушное охлаждение	RSC ScaleStream реализация	RSC ScaleStream предельно (40U)	Примечания
Сервер 4U: 2 ЦПУ 8 ускорителей	3000 Вт	3000 Вт	3000 Вт	Без учета снижения потребления за счет вентиляторов
Охлаждается жидкостью / сервер	0 Вт	2690 Вт	2690 Вт	Ускорители $295 \cdot 8 = 2360$ ЦПУ $165 \cdot 2 = 330$
Охлаждается воздухом / сервер	3000 Вт	310 Вт	310 Вт	
Предельная мощность воздушного охлаждения на стойку	6000 Вт	6000 Вт	6000 Вт	Ограничение
Кол-во серверов в стойке	2	7	10	Стойка 42 U не более 10 серверов
Мощность потребления на стойку	6000 Вт	21000 Вт	30000 Вт	
Охлаждается жидкостью / стойку	0 Вт	18830 Вт	26900 Вт	
Охлаждается воздухом / стойку	6000 Вт	2170 Вт	3100 Вт	Воздушное охлаждение не является ограничением. 20 установленных платформ занимают 3 стойки вместо 10

Охлаждение процессоров AMD, Intel Xeon Gen4 и ускорителей A100



Что может дать добавление жидкости?

ASMB10-iKVM

Firmware Information
1.2.21
Jan 3 2023 17:12:04 UTC
Host Online

Quick Links..

- Dashboard
- Sensor
- System Inventory
- FRU Information
- Logs & Reports

System Inventory

Processor Memory Controller BaseBoard Power PCIE Device Storage GPU Card

Processor Info

Id	Manufacturer	Brand Name	State	MaxSpeedMHz	TotalCores
CPU0	Intel	Intel(R) Xeon(R) Gold 6342 CPU @ 2.80GHz	Enabled	4000	24
CPU1	Intel	Intel(R) Xeon(R) Gold 6342 CPU @ 2.80GHz	Enabled	4000	24

ASMB10-iKVM

Firmware Information
1.2.21
Jan 3 2023 17:12:04 UTC
Host Online

Quick Links..

- Dashboard
- Sensor
- System Inventory
- FRU Information
- Logs & Reports

System Inventory

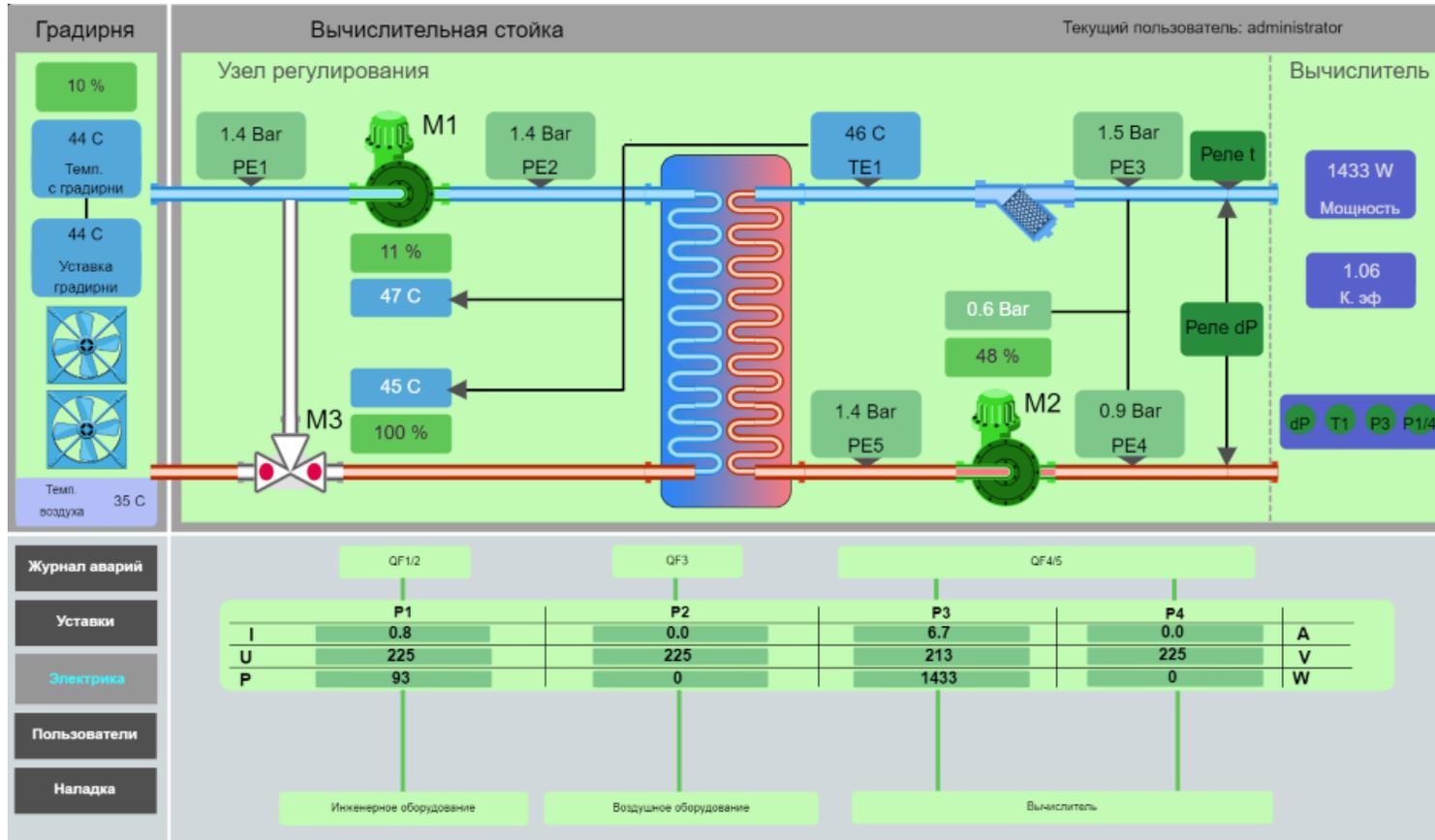
Processor Memory Controller BaseBoard Power PCIE Device Storage GPU Card

temperature event occurred

GPU Information

Model Name	Serial No.	FW Version	Current Temperature	Current Watts(W)
NVIDIA A100 80GB PCIe	1654822654218	92.00.9A.00.01	84	301
NVIDIA A100 80GB PCIe	1654822654176	92.00.9A.00.01	77	309

Инфраструктура эксперимента



Что будет, если нагрузить по полной?

```
root@gfxnode:~
Every 1.0s: nvidia-smi
Wed Oct 25 13:52:40 2023
+-----+
| NVIDIA-SMI 545.23.06                Driver Version: 545.23.06 |
+-----+-----+
| GPU  Name      Persistence-M | Bus-Id  Persistence-M |
| Fan  Temp     Perf          Pwr:Usage/Cap |   Vbios Part   Pwr:Usage/Cap |
+-----+-----+-----+-----+
|  0  NVIDIA A100 80GB PCIe      Off          | 00000000  Off          |
| N/A   77C      P0              301W / 300W | 72887MiB / 79680MiB |
+-----+-----+-----+-----+
|  1  NVIDIA A100 80GB PCIe      Off          | 00000000  Off          |
| N/A   84C      P0              306W / 300W | 72887MiB / 79680MiB |
+-----+-----+-----+-----+
+-----+
| Processes: |
| GPU  GI    CI       PID  Type  Process name      GPU Memory |
| ID   ID   ID           |          |                  | Usage     |
+-----+-----+-----+-----+
|  0   N/A  N/A         3481  C    ./gpu_burn        72874MiB |
|  1   N/A  N/A         3506  C    ./gpu_burn        72874MiB |
+-----+-----+-----+-----+
DGEMM 1 in 26.598 sec| 3119.76 GFLOPS | CPU_Freq 2398-2398-2399,2747-2748-2748 MHz| UNC_Freq 1267,1460 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.3,33.4 W| CPU_Temp 89, 75 °C| 2023-10-25T13:45:26.511-04:00| Limits PAC,PAC
DGEMM 1 in 26.400 sec| 3143.12 GFLOPS | CPU_Freq 2387-2388-2388,2734-2735-2735 MHz| UNC_Freq 1289,1490 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.5,33.6 W| CPU_Temp 89, 75 °C| 2023-10-25T13:45:52.913-04:00| Limits PAC,PAC
DGEMM 1 in 26.500 sec| 3131.33 GFLOPS | CPU_Freq 2392-2392-2393,2739-2740-2740 MHz| UNC_Freq 1275,1471 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.4,33.6 W| CPU_Temp 81, 75 °C| 2023-10-25T13:46:19.414-04:00| Limits PAC,PAC
DGEMM 1 in 26.570 sec| 3123.02 GFLOPS | CPU_Freq 2392-2393-2393,2733-2734-2734 MHz| UNC_Freq 1304,1506 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.4,33.5 W| CPU_Temp 89, 75 °C| 2023-10-25T13:46:45.986-04:00| Limits PAC,PAC
DGEMM 1 in 26.418 sec| 3141.03 GFLOPS | CPU_Freq 2387-2388-2388,2736-2737-2737 MHz| UNC_Freq 1288,1486 MHz| CPU_Pwr 229
.3,229.0 W| DRAM_Pwr 33.6,33.8 W| CPU_Temp 86, 75 °C| 2023-10-25T13:47:12.406-04:00| Limits PAC,PAC
DGEMM 1 in 26.408 sec| 3142.25 GFLOPS | CPU_Freq 2397-2398-2398,2743-2744-2745 MHz| UNC_Freq 1262,1444 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.3,33.5 W| CPU_Temp 92, 76 °C| 2023-10-25T13:47:38.815-04:00| Limits PAC,PAC
DGEMM 1 in 26.441 sec| 3138.32 GFLOPS | CPU_Freq 2393-2394-2394,2738-2739-2740 MHz| UNC_Freq 1268,1473 MHz| CPU_Pwr 229
.3,229.0 W| DRAM_Pwr 33.5,33.7 W| CPU_Temp 82, 75 °C| 2023-10-25T13:48:05.257-04:00| Limits PAC,PAC
DGEMM 1 in 26.357 sec| 3148.35 GFLOPS | CPU_Freq 2395-2397-2397,2740-2740-2740 MHz| UNC_Freq 1274,1459 MHz| CPU_Pwr 229
.4,229.0 W| DRAM_Pwr 33.4,33.5 W| CPU_Temp 82, 75 °C| 2023-10-25T13:48:31.615-04:00| Limits PAC,PAC
DGEMM 1 in 26.225 sec| 3164.14 GFLOPS | CPU_Freq 2387-2387-2387,2729-2730-2730 MHz| UNC_Freq 1293,1486 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.4,33.5 W| CPU_Temp 94, 75 °C| 2023-10-25T13:48:57.842-04:00| Limits PAC,PAC
DGEMM 1 in 26.580 sec| 3121.93 GFLOPS | CPU_Freq 2397-2398-2398,2734-2735-2736 MHz| UNC_Freq 1298,1492 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.5,33.6 W| CPU_Temp 82, 76 °C| 2023-10-25T13:49:24.423-04:00| Limits PAC,PAC
DGEMM 1 in 26.320 sec| 3151.01 GFLOPS | CPU_Freq 2392-2392-2392,2726-2728-2728 MHz| UNC_Freq 1309,1508 MHz| CPU_Pwr 229
.4,229.0 W| DRAM_Pwr 33.6,33.7 W| CPU_Temp 84, 76 °C| 2023-10-25T13:49:50.752-04:00| Limits PAC,PAC
DGEMM 1 in 26.559 sec| 3150.42 GFLOPS | CPU_Freq 2397-2398-2398,2737-2740-2740 MHz| UNC_Freq 1276,1463 MHz| CPU_Pwr 229
.3,229.0 W| DRAM_Pwr 33.4,33.6 W| CPU_Temp 78, 76 °C| 2023-10-25T13:50:17.093-04:00| Limits PAC,PAC
DGEMM 1 in 26.326 sec| 3152.04 GFLOPS | CPU_Freq 2398-2399-2399,2737-2739-2739 MHz| UNC_Freq 1267,1457 MHz| CPU_Pwr 229
.3,229.0 W| DRAM_Pwr 33.6,33.7 W| CPU_Temp 86, 76 °C| 2023-10-25T13:50:43.420-04:00| Limits PAC,PAC
DGEMM 1 in 26.443 sec| 3138.06 GFLOPS | CPU_Freq 2401-2402-2403,2734-2735-2736 MHz| UNC_Freq 1292,1486 MHz| CPU_Pwr 229
.3,228.9 W| DRAM_Pwr 33.5,33.5 W| CPU_Temp 87, 77 °C| 2023-10-25T13:51:09.865-04:00| Limits PAC,PAC
```



CPU1 Temperature	83 °C
CPU2 Temperature	77 °C
CPU_Power	456 Watts

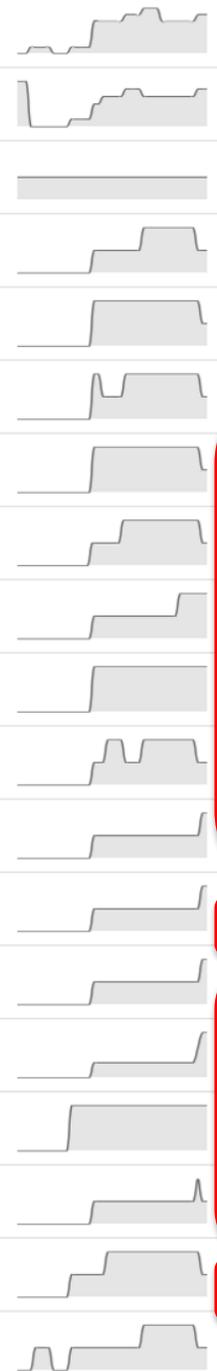
DIMMA1_Temp	62 °C
DIMMB1_Temp	59 °C
DIMMC1_Temp	58 °C
DIMMD1_Temp	62 °C

DIMME1_Temp	64 °C
-------------	-------

DIMMF1_Temp	61 °C
DIMMG1_Temp	60 °C
DIMMH1_Temp	63 °C
DIMMJ1_Temp	56 °C

DIMMK1_Temp	54 °C
DIMML1_Temp	56 °C
DIMMM1_Temp	58 °C
DIMMN1_Temp	53 °C

DIMMP1_Temp	51 °C
DIMMR1_Temp	51 °C
DIMMT1_Temp	53 °C



Потребление сервера по 220В	1408 Вт
Потребление сервера по 12В	1328 Вт
Потребление блоков питания	80 Вт
Потребление памяти	64 Вт
Потребление процессоров	456 Вт
Потребление GPU	600 Вт
Остальная система и вентиляторы	208 Вт
ИТОГО воздух	352 Вт
ИТОГО жидкость	1056 Вт

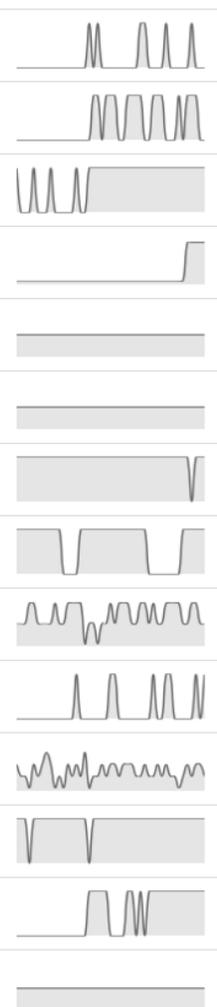
FRNT_FAN1	3360 RPM
FRNT_FAN2	4920 RPM
FRNT_FAN3	5040 RPM
FRNT_FAN4	4920 RPM
FRNT_FAN5	3360 RPM
FRNT_FAN6	3360 RPM
FRNT_FAN7	3360 RPM

Memory_Power	64 Watts
--------------	----------

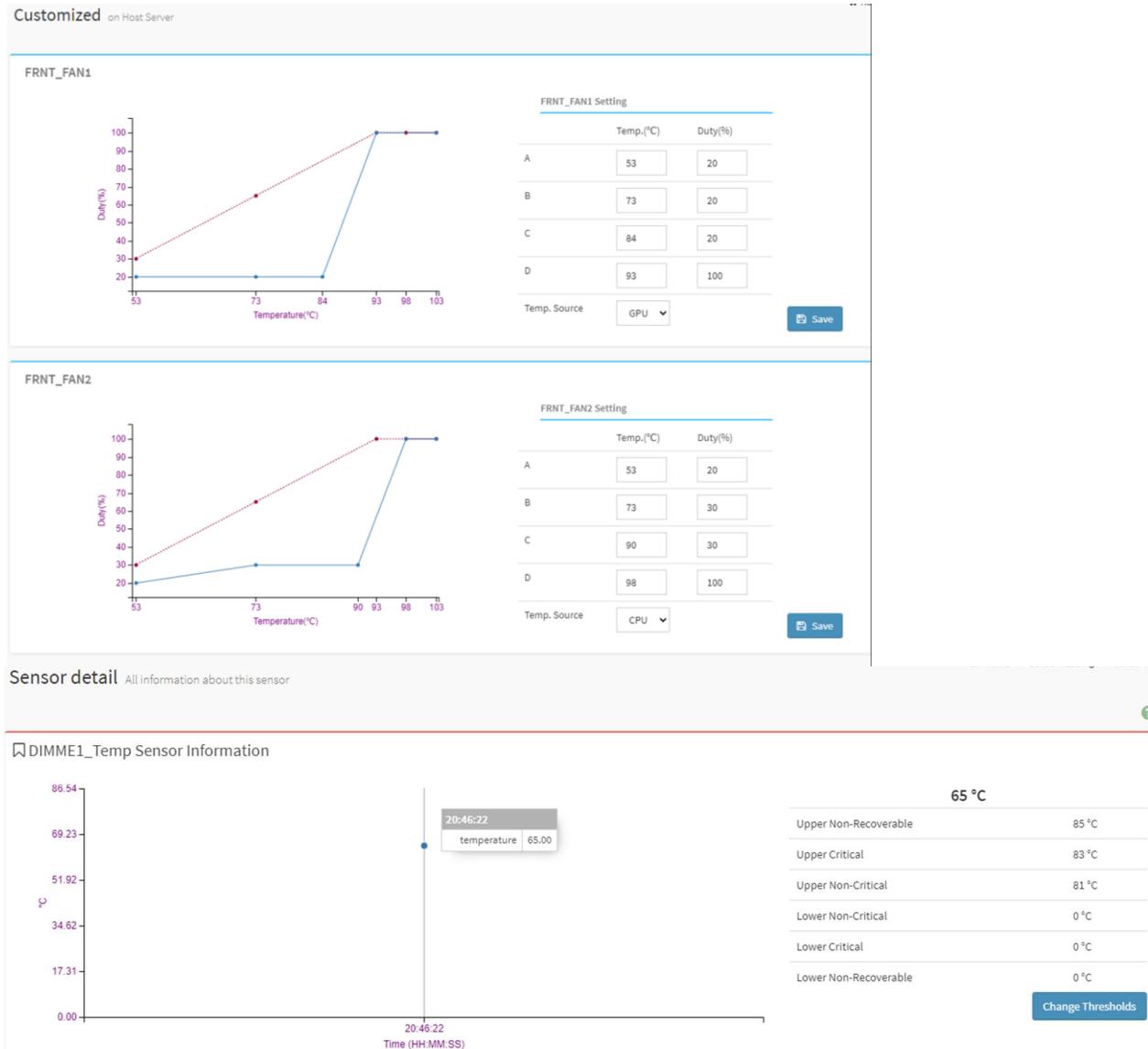
PSU1 Power In	688 Watts
PSU1 Power Out	656 Watts
PSU2 Power In	720 Watts
PSU2 Power Out	672 Watts

TR1 Temperature	33 °C
-----------------	-------

VBAT	3.12 Volts
------	------------



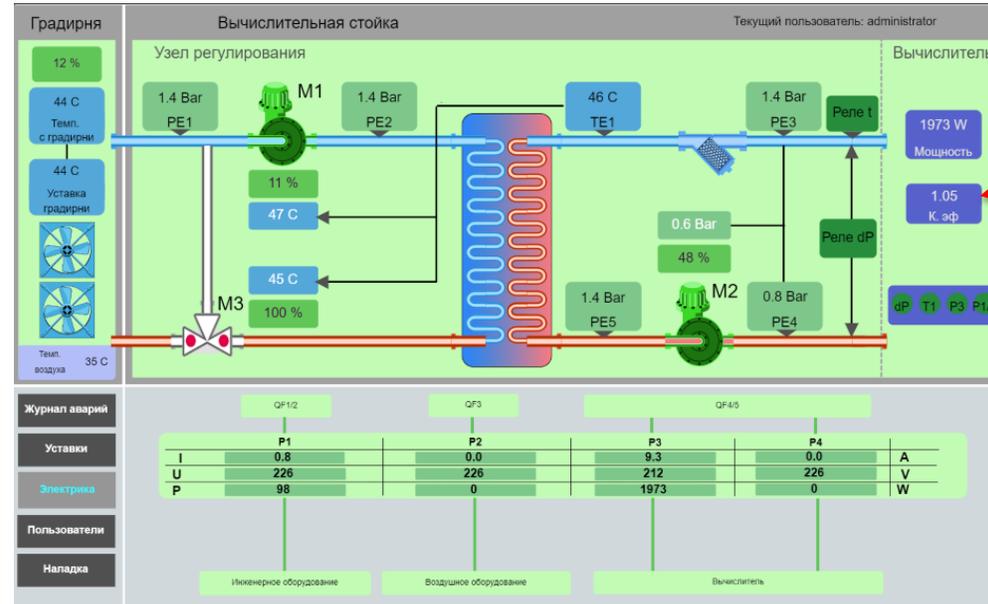
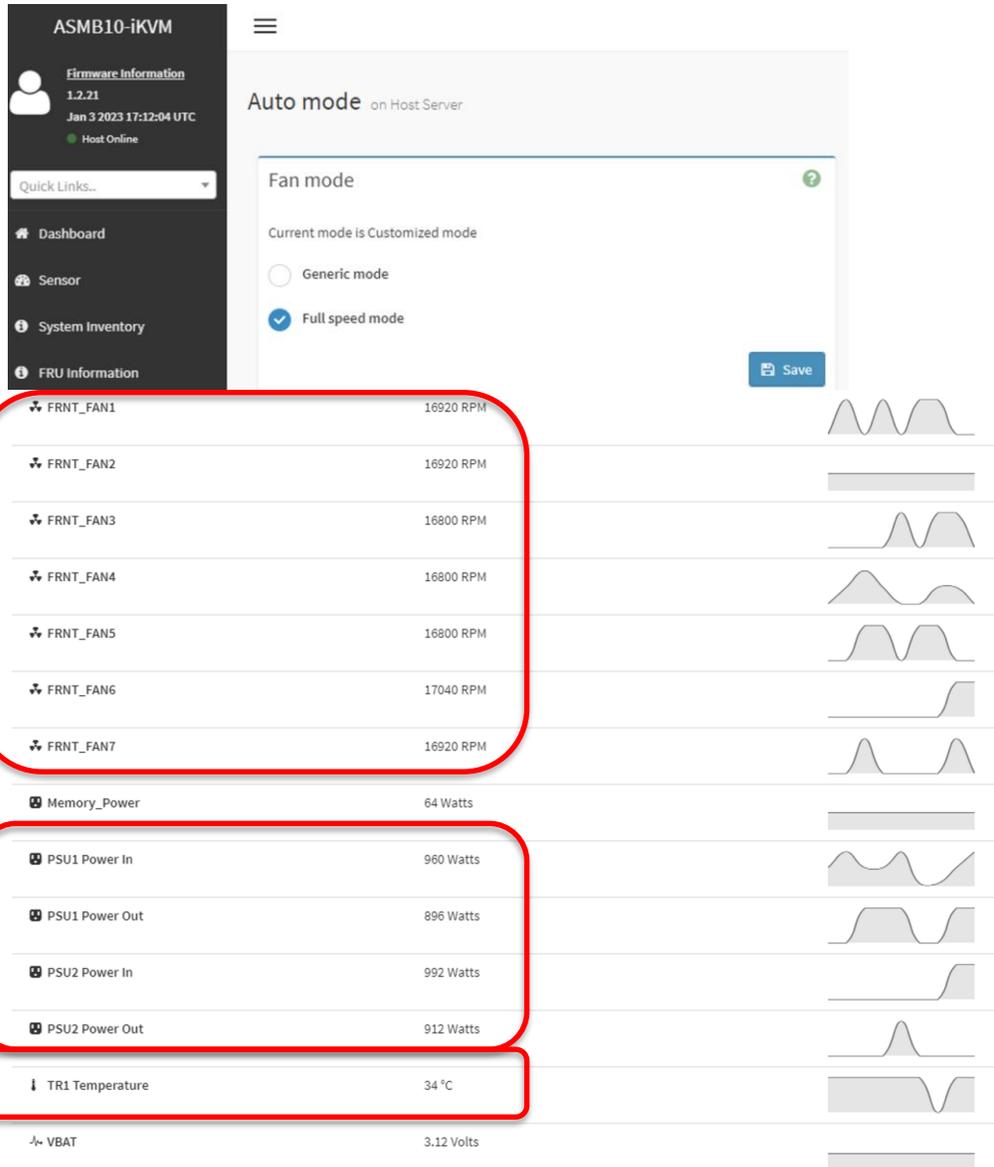
Профиль оборудования важен



Необходимо понимание каким образом охлаждаются компоненты:

- На вентиляторах GPU остаются лишь цепи питания
- На вентиляторах CPU остаются VRM и внешние сетевые адаптеры
- Память имеет низкую энергетическую плотность и низкие требования к охлаждению
- Диски SSD не имеют проблем с охлаждением

А если только воздух?



PUE стал лучше?
Причины?
Вентиляторы в сервере
входят в ИТ нагрузку!
Но это же система
охлаждения!!!

$$(1973+98)/1973=1,05$$

$$(1973+98)/1473=1,41$$

Расчетная нагрузка	Вт/сервер	6 кВт/стойку	10 кВт/стойку
Кол-во серверов потребляющих по 220В	1 952	3	5
ИТОГО воздух	1 952	5 856	9 760
Затраты на воздушное охлаждение (0.3)	586	1 757	2 928
Общее потребление		7 613	12 688

Жидкость поможет!

Расчетная нагрузка	Вт/сервер	6 кВт/стойку	10 кВт/стойку
Кол-во серверов потребляющих по 220В	1 408	4	7
ИТОГО воздух	352	1 408	2 464
ИТОГО жидкость	1 056	4 224	7 392
Затраты на воздушное охлаждение (0.3)	106	422	739
Затраты на жидкостное охлаждение (0.1)	106	422	739
Общее потребление		6 477	11 334



Компонуемая архитектура «PCK Торнадо»
на базе Intel Xeon Scalable 3-го поколения

967,45 ТФлопс**

Вычислительная плотность
на шкаф*

3,67 ТБ/с

Пропускная способность
распределенной системы
хранения

604,66 ТФлопс/м³

Производительность на объем

130 кВт

Энергетическая
плотность на шкаф



* шкаф 42U 80x100 см

** для решений на базе Intel® Xeon®

Деагрегированная компонентная инфраструктура



Вычислительные узлы

с поддержкой процессоров Intel, AMD и «Эльбрус»



Гиперконвергентные узлы

с устройствами хранения NVMe

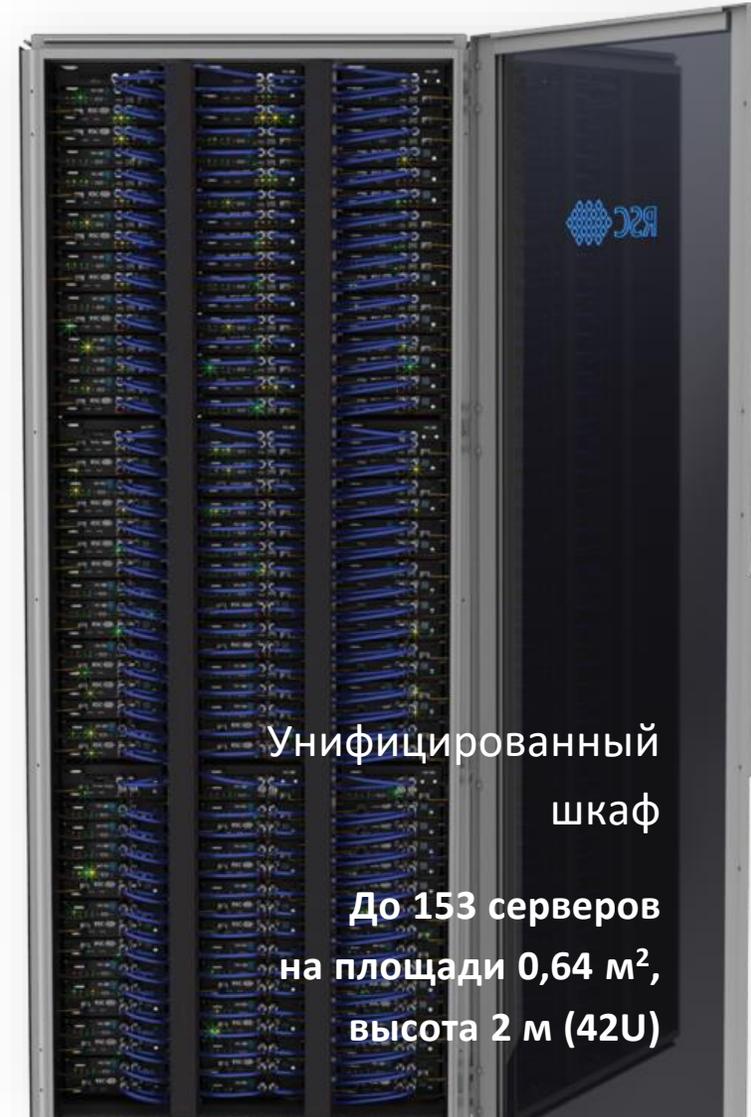


Модули избыточного питания



Программный стек управления

RSC Basis Software Platform



Унифицированный
шкаф

До 153 серверов
на площади 0,64 м²,
высота 2 м (42U)

Вычислительный узел «РСК Торнадо» на базе Intel Xeon Scalable 3-го поколения



- **100% охлаждение «горячей водой»** позволяет достичь рекордной энергоэффективности (PUE < 1,04)
- **2x процессора Intel® Xeon® Scalable 3-го поколения:** Platinum 83xx (TDP 270W) или Gold 63xx (TDP 205W)
- **8 модулей энергонезависимой памяти Intel® Optane DC Persistent Memory (до 4 ТБ)**
- **5 NVMe SSD** диска E1.S/M.2 (2 Hot Swap) до 16 ТБ хранения
- Intel® Omni-Path 100 Gb/s, Infiniband 100/200Gb/s, High speed Ethernet
- 2 порта 10Gb/s Ethernet

СХД PCK Tornado AFS рекордно большого объема



E1.L Intel® Data Center SSDs
в форм-факторе EDSFF



Сверхвысокая емкость – 1 ПБ
в одном сервере (1U)



Надежное объединение 2 серверов
в СХД емкостью 2 ПБ (2U)

Первое решение на 100% жидкостном
охлаждении с высочайшей
плотностью на базе 32x Intel EDSFF
SSDs, двух процессоров Intel® Xeon®
Scalable и памяти Intel® Optane™ DC
Persistent Memory

**100% охлаждение «горячей
водой» позволяет достичь
рекордной энергоэффективности
(CUE < 1,04)**



РСК Торнадо AFS с новым Intel SSD P5316

41.3 ПБ

Рекордная емкость
теплой системы
хранения на шкаф

1 ТБ/с

Пропускная способность
распределенной системы
хранения

50 кВт

Энергетическая
плотность на шкаф



шкаф 42U 60x120 см

Система хранения «по запросу»

«PCK БазИС» позволяет создавать системы хранения данных «по запросу»:

Кластерная файловая система Lustre

Стандарт «де-факто» в мире суперкомпьютеров

Новая высокопроизводительная объектная система хранения DAOS

Разработана «с чистого листа» для поддержки высокоскоростных фабрик, устройств NVMe и Storage Class Memory

Предоставляет современные высокопроизводительные методы работы с данными:

HDFS

Apache Spark

MPI-IO

TensorFlow

NoSQL

S3

POSIX

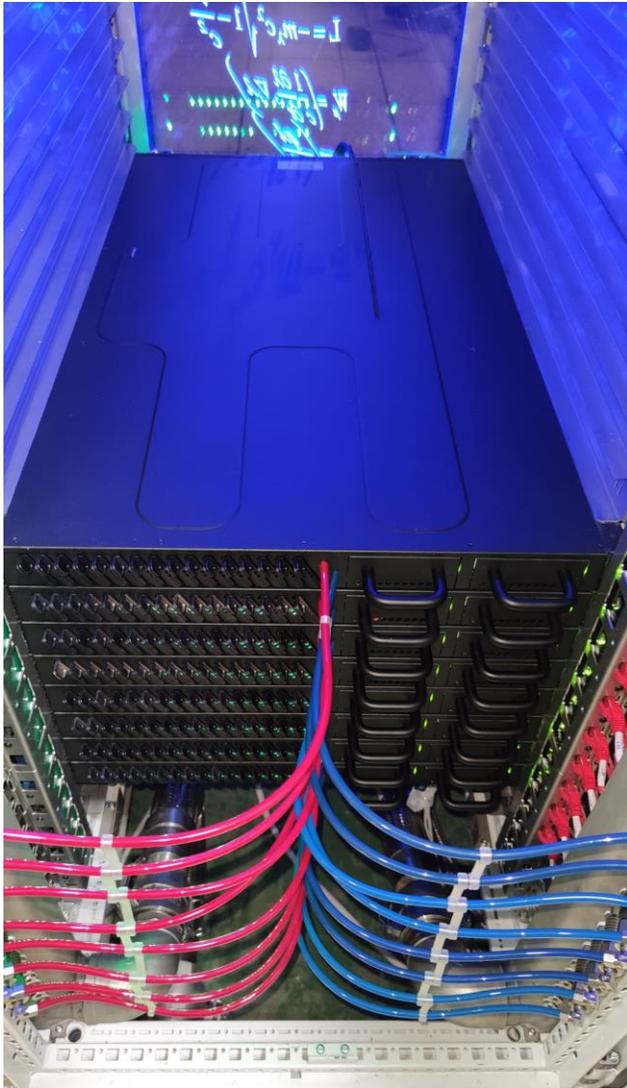


Система хранения «по запросу» PCK БазИС



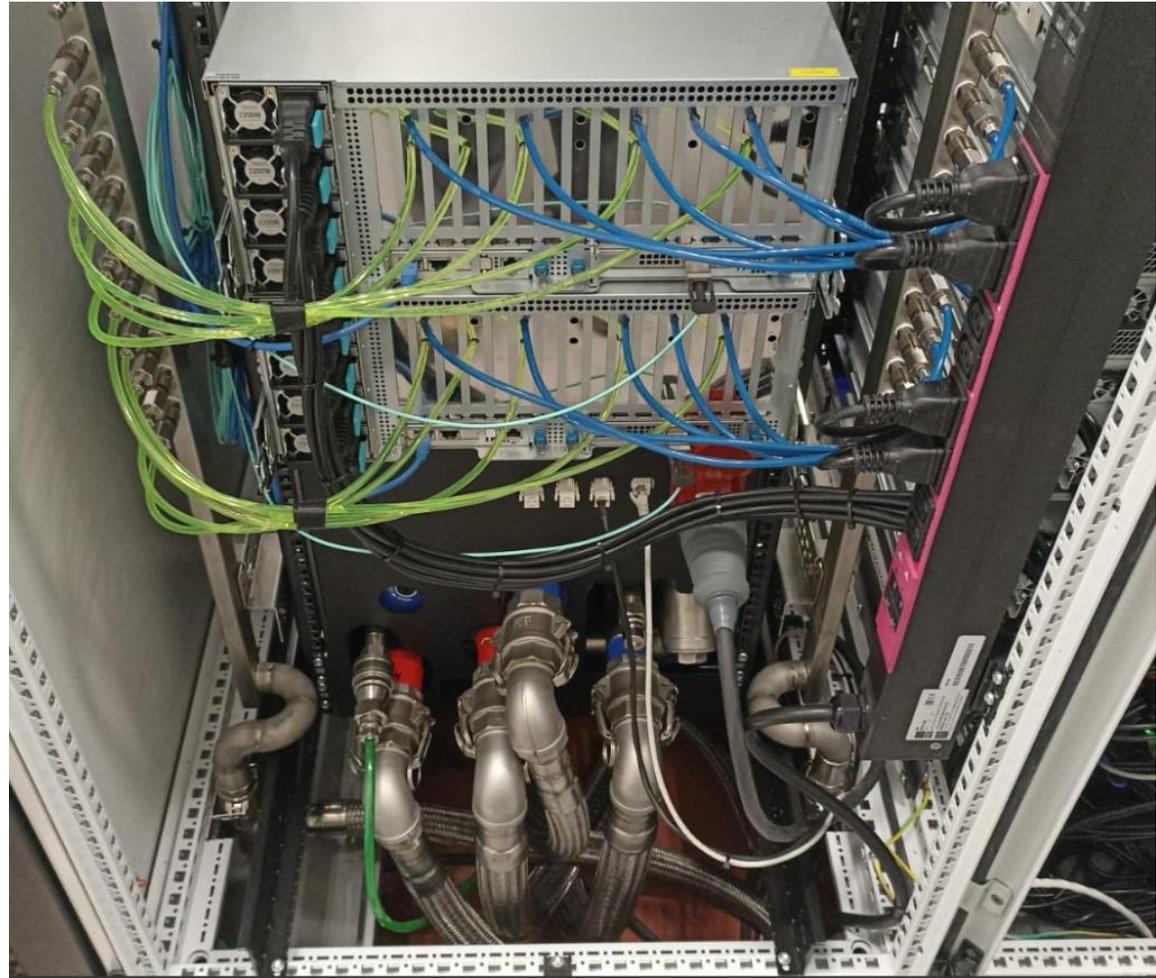
Группа компаний PCK получила престижную награду Russian DC Awards 2020 в номинации «Лучшее ИТ-решение для ЦОДа», победив с проектом «Высокопроизводительная система хранения для суперкомпьютера», реализованном в 2020 году в Объединенном институте ядерных исследований (ОИЯИ) в Дубне.

СХД 8ПБ в ОИЯИ



МикроЦОД





Реализованные проекты



Южно-Уральский государственный университет в 2010 году. В 2013 году после модернизации занимал 127-е место в списке Top500 (ноябрь 2013 г).



Межведомственный суперкомпьютерный центр РАН

Реализованные проекты

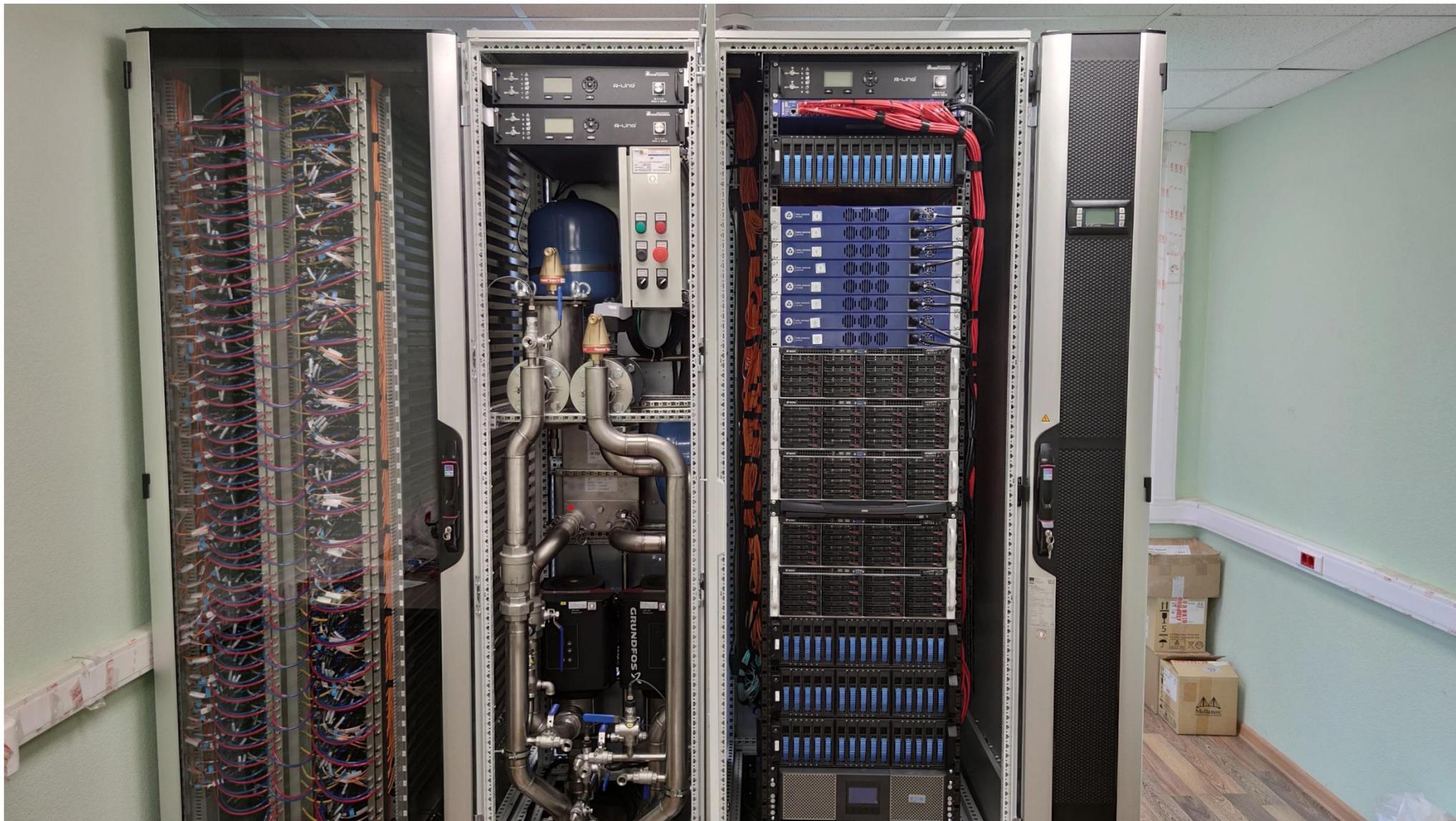


Объединенный институт ядерных исследований (ОИЯИ), Дубна

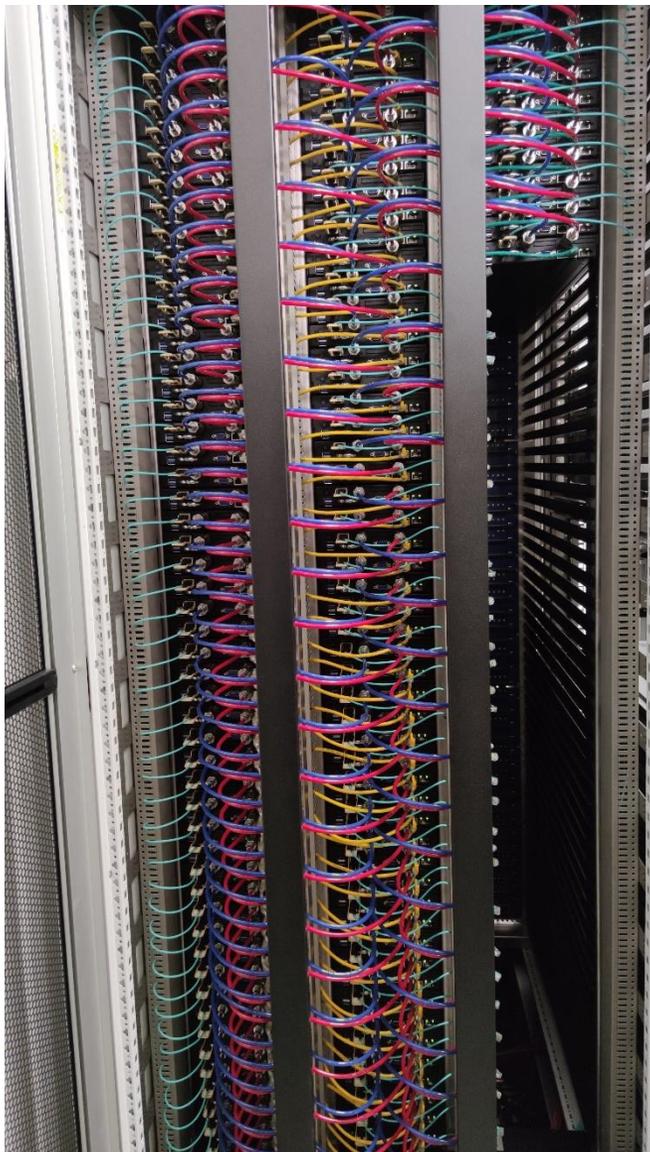


Санкт-Петербургский политехнический университет имени Петра Великого (СПбПУ)

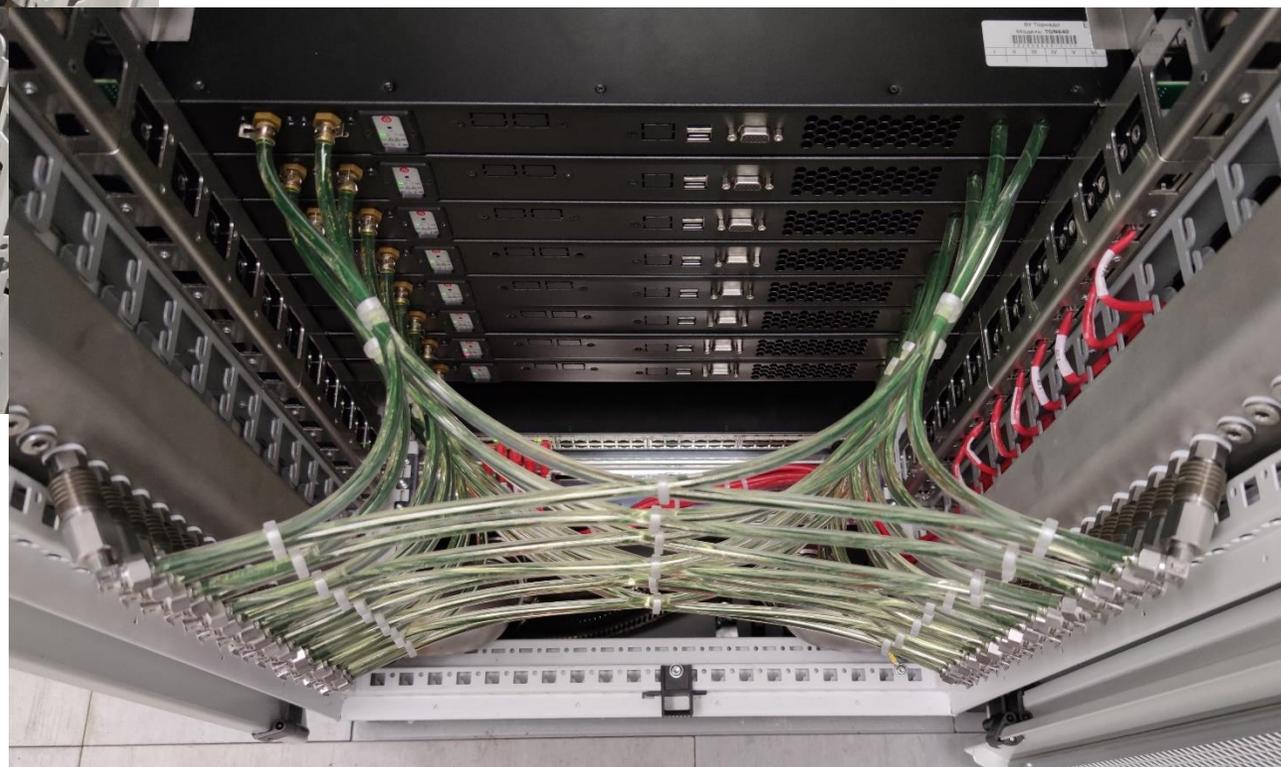
Реализованные проекты. Автономный ЦОД ЦИАМ.



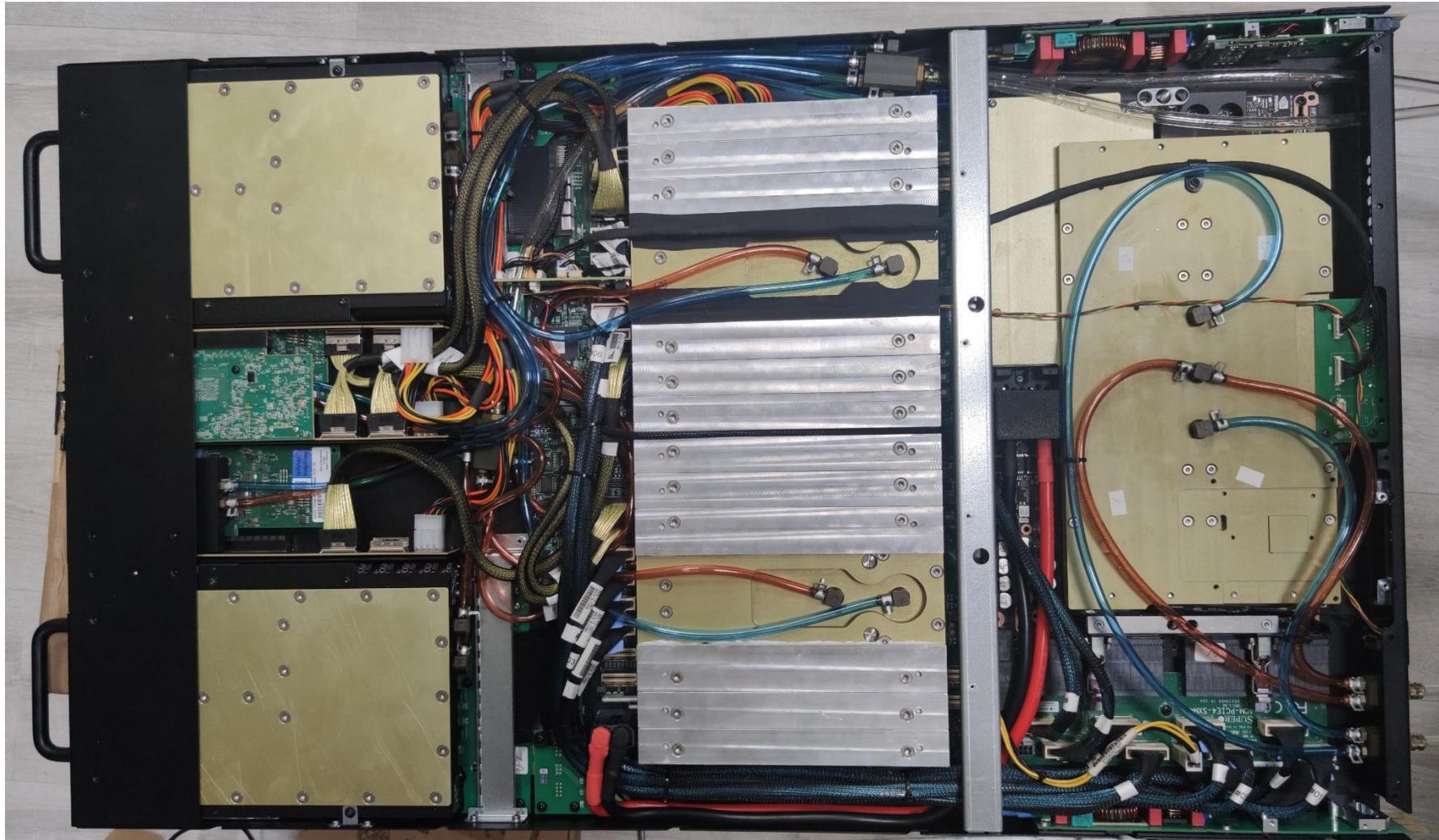
Вычислитель, ЦИАМ.



Сервера с ускорителями А100, ЦИАМ.



Сервер с ускорителями 4 x A100 в формате SXM



Спасибо!



rscgroup.ru

migal@rsc-systems.ru