



Железный фундамент

Как автоматизация помогает строить и динамично менять инфраструктуру облака

Антон Бушный

Старший инженер,
Infrastructure Team lead

Про что поговорим?

- Про железо в комплексе: без сравнений, без замеров производительности
- Про подходы к проектированию и управлению большой инфраструктурой
- Про то как можно упростить работу с помощью автоматизаций

Наша инфраструктура в цифрах



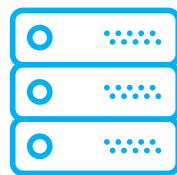
5

ЦОДы
присутствия



>110

Стойки



>1600

Серверы



>450

Сетевое
оборудование



>450

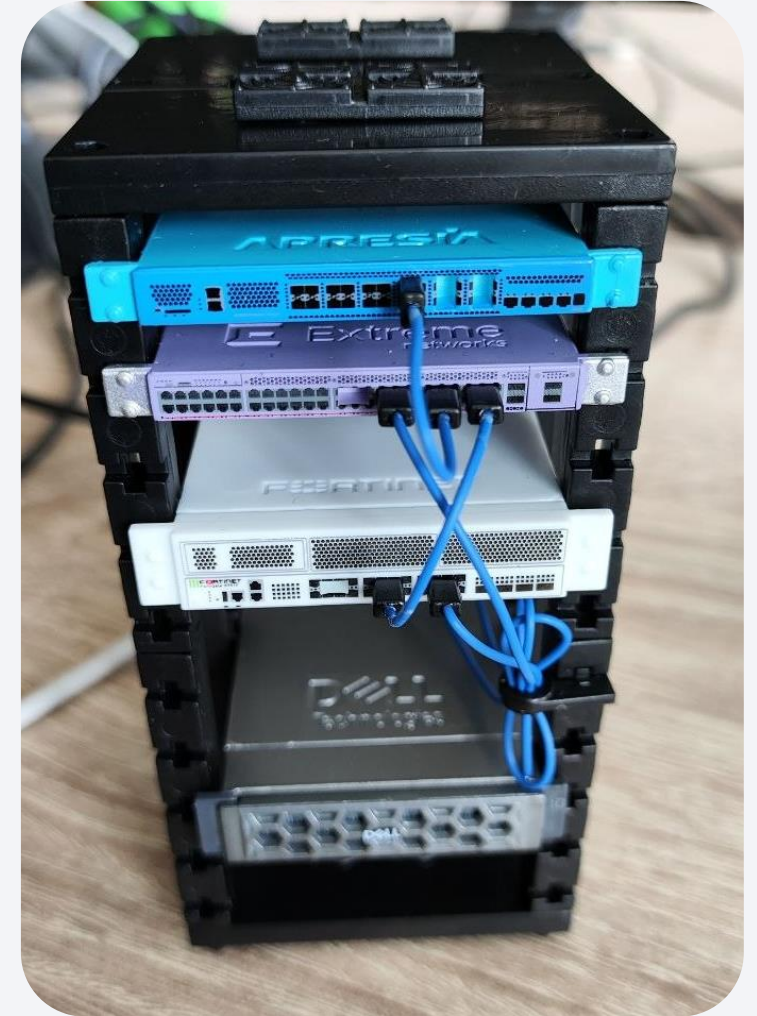
Прочее
оборудование



+ частные инсталляции платформы

Наша инфраструктура динамичная

- Регулярно вводится новое оборудование, выводится старое
- Оборудование часто перемещается между ЦОД
- Оборудование меняет владельцев (например может сдаваться в аренду)



Что-то может пойти не по плану всегда



Мы с утра приехали, а тут вот

ахахахах



Вон там торчит флаг цода

Как было раньше ≈ 2018-2019



Network



Storage



Compute

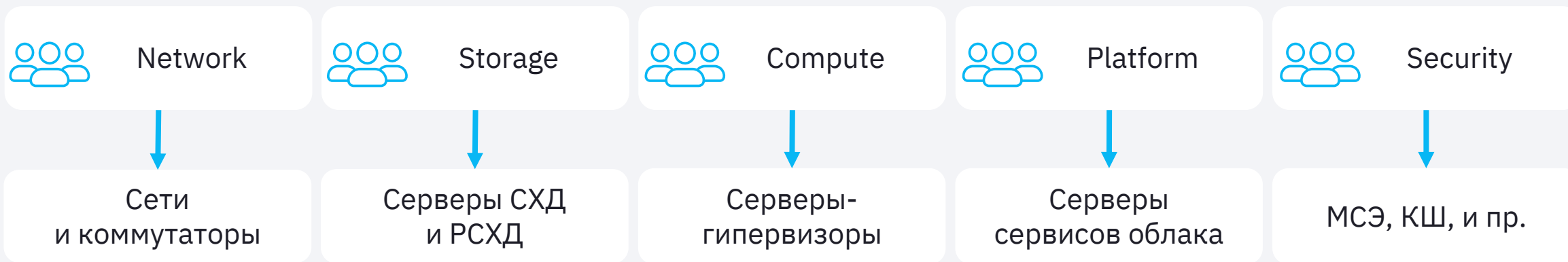


Platform



Security

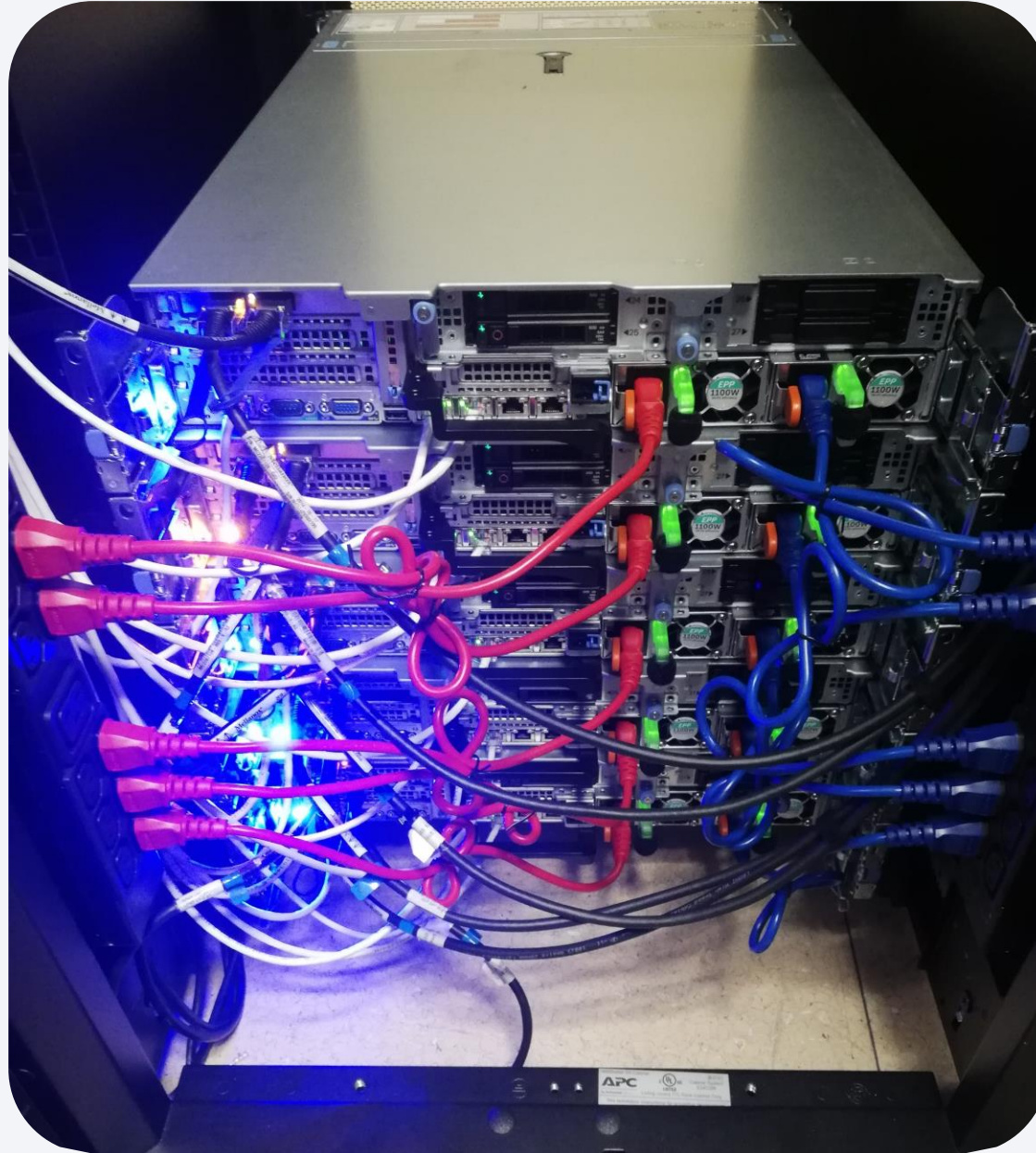
Как было раньше ≈ 2018-2019



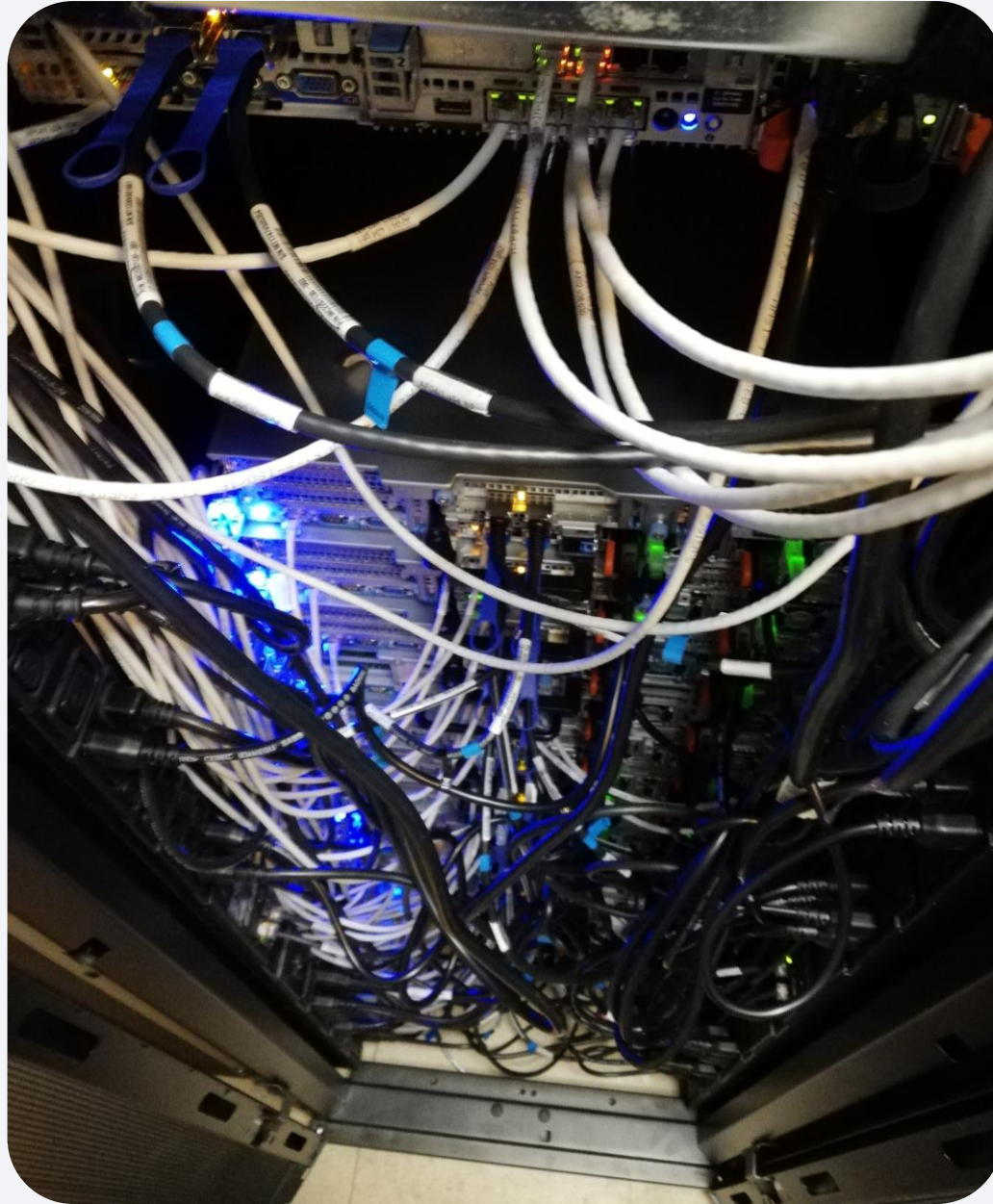
Как было раньше ≈ 2018-2019



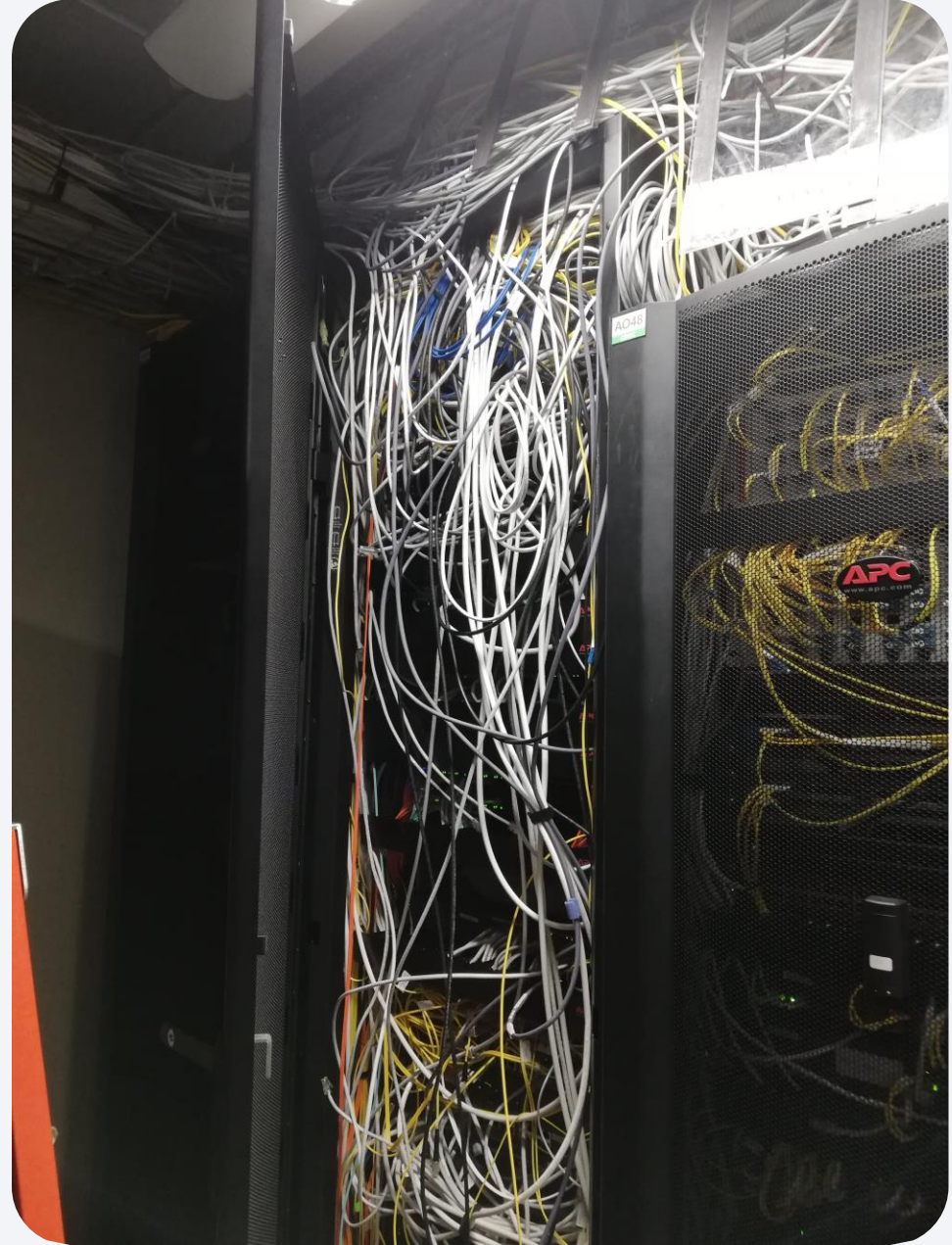
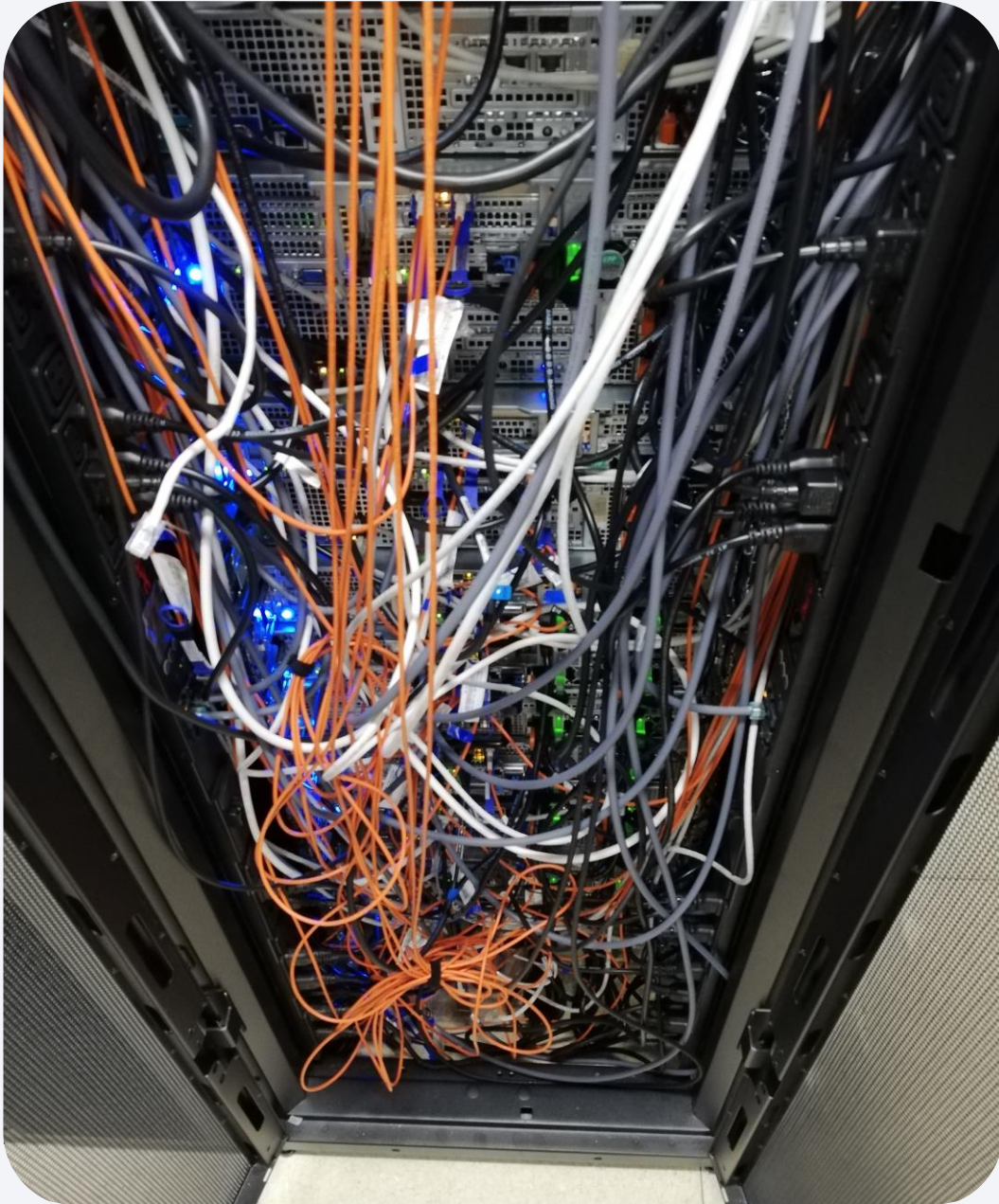
Что из этого получалось....



Что из этого получалось спустя месяц...

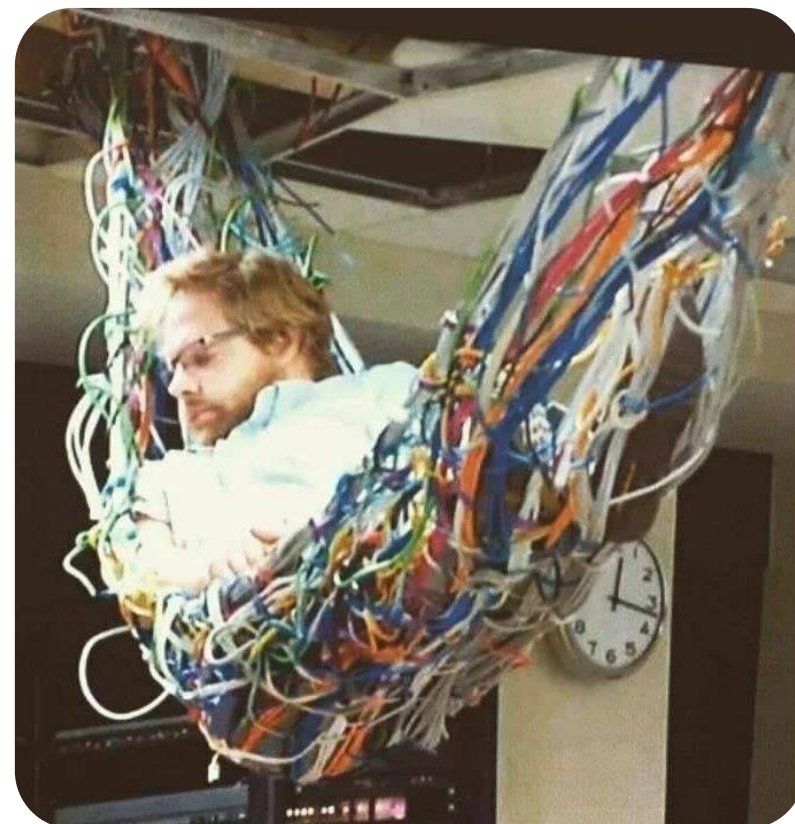
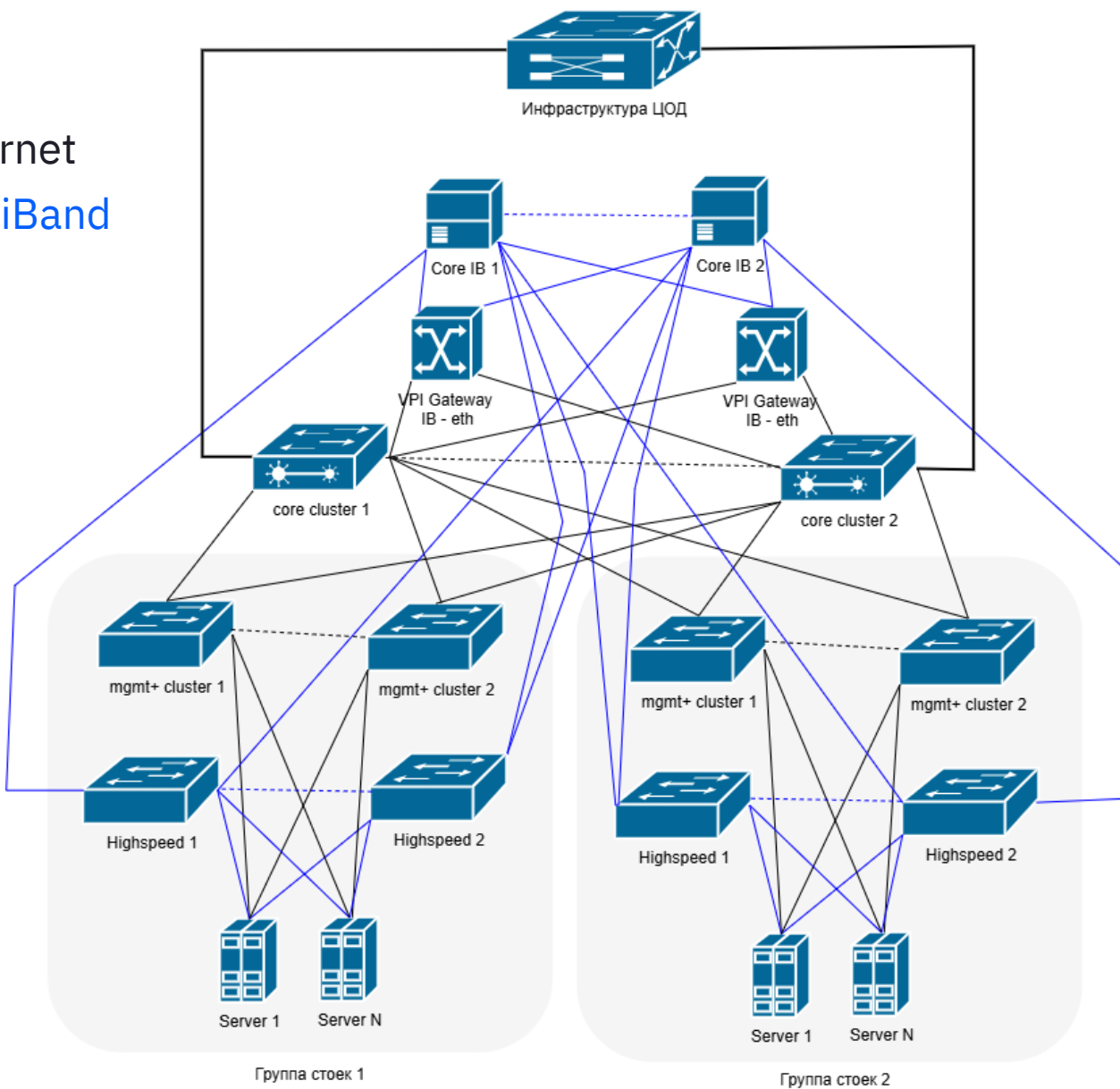


Что из этого получалось спустя несколько месяцев



Старая сеть

- Ethernet
- InfiniBand

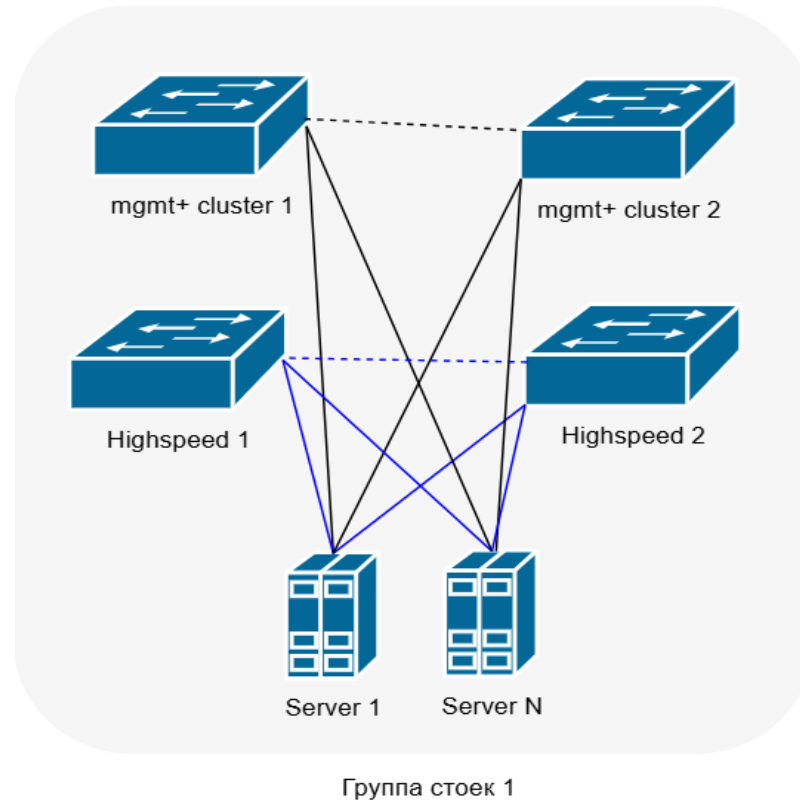


Старая сеть

- Ethernet
- InfiniBand

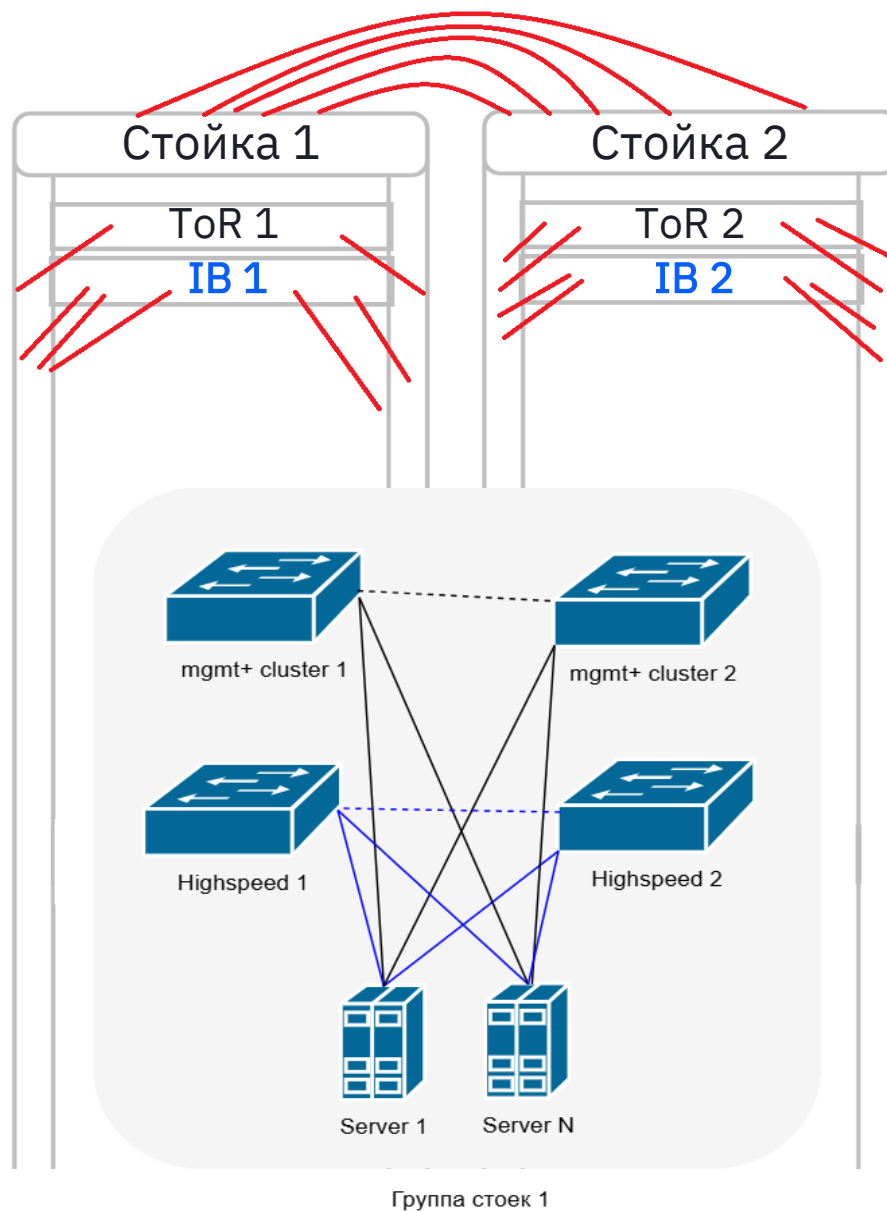
Стойка 1

Стойка 2



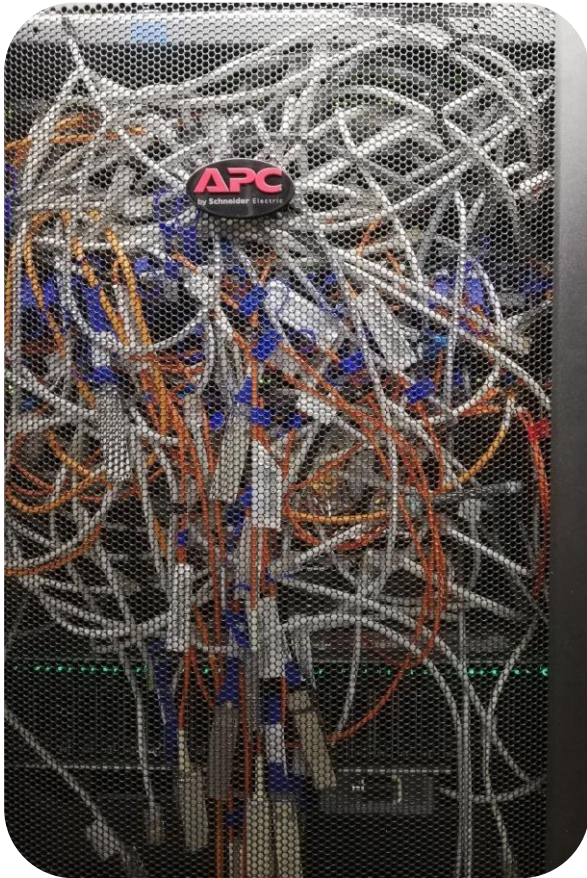
Старая сеть

- Ethernet
- InfiniBand

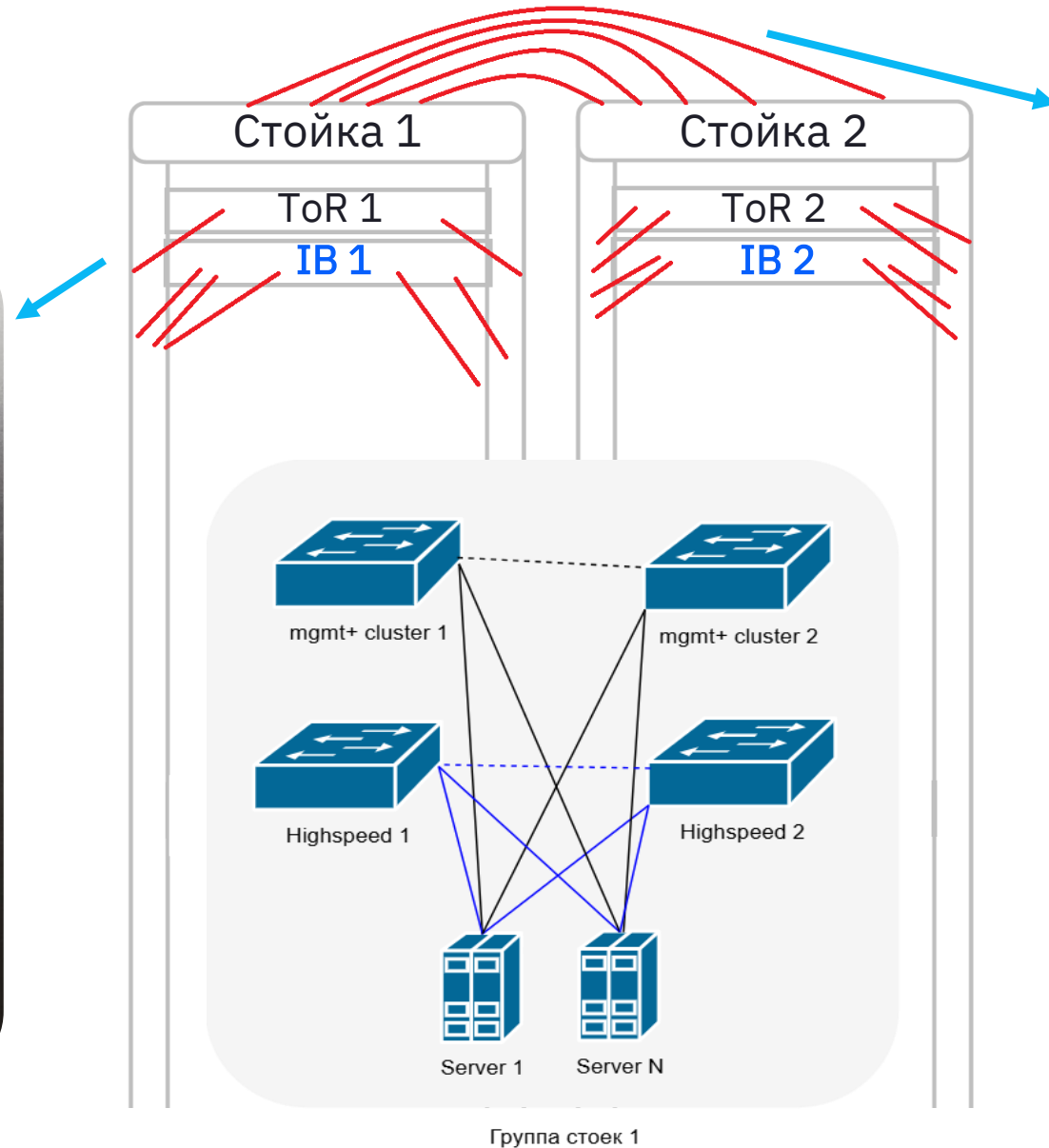


Старая сеть

- Ethernet
- InfiniBand



K2 CLOUD



А минусы будут?



А минусы будут?



Частые сбои при проведении работ в ЦОД

А минусы будут?



Частые сбои при проведении работ в ЦОД



Несбалансированная заполненность стоек



Неравномерная нагрузка по электропитанию

А минусы будут?



Частые сбои при проведении работ в ЦОД



Несбалансированная заполненность стоек



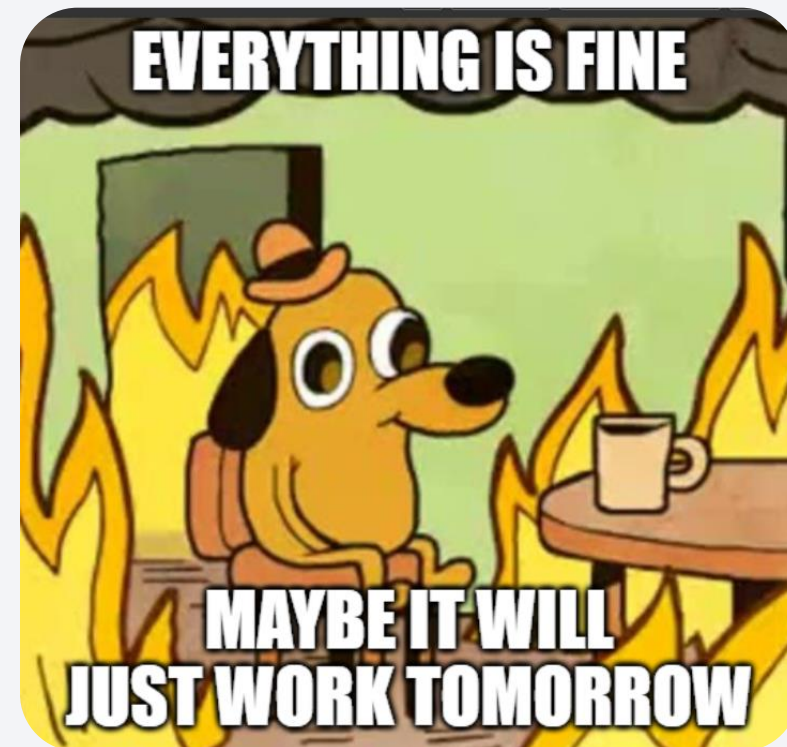
Неравномерная нагрузка по электропитанию



Негатив от части инженеров

А минусы будут?

- ! Частые сбои при проведении работ в ЦОД
- ! Несбалансированная заполненность стоек
- ! Неравномерная нагрузка по электропитанию
- ! Негатив от части инженеров





Турбулентность ($\approx$ 2019-2020)

Турбулентность

- ✗ Несколько стоек полностью отключились из-за перегрузки по питанию
- ✗ Сеть устарела, приходилось перезагружать хосты
- ✗ «Отказоустойчивые» СХД падали друг за другом вместе с данными



K2 CLOUD

```
Resolved in 47m 35s: IB queue stopped
Problem has been resolved in 47m 35s at 12:57:30 on 2022.02.28
Problem name: IB queue stopped
Host: GrayHook
Severity: Disaster
Original problem ID: 74781281
```

15:57

```
Problem: IB queue stopped
Problem started at 12:57:55 on 2022.02.28
Problem name: IB queue stopped
Host: GrayHook
Severity: Disaster
Operational data [REDACTED], message ib0: queue (5)
stopped, tx-head 450398777, tx-tail 450395694
Original problem ID: 74793692
```

15:57



```
Error: V-35-410: Cluster Server not running on local node: mg-a
- Mount points... not ok
* ERROR : CFS is not mounted.

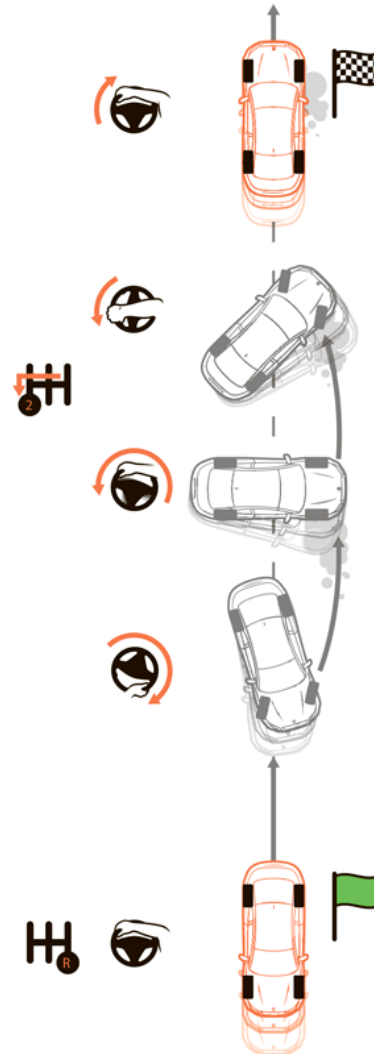
- I/O Fencing status
FEN vxfenadm ERROR V-11-2-1115 Local node is not a member of cluster!
- Fencing mode ...
t ok
* ERROR : I/O fencing mode is not set to violin.

- NTB status ...
- vntb0 configured ... ok
- vntb0 link status ... ok

Summary : Found 3 errors.
lp-violin-05-mg-a [comlp-violin-05: master] #
```

Нужно всё менять! И довольно быстро...

29. Полицейский разворот



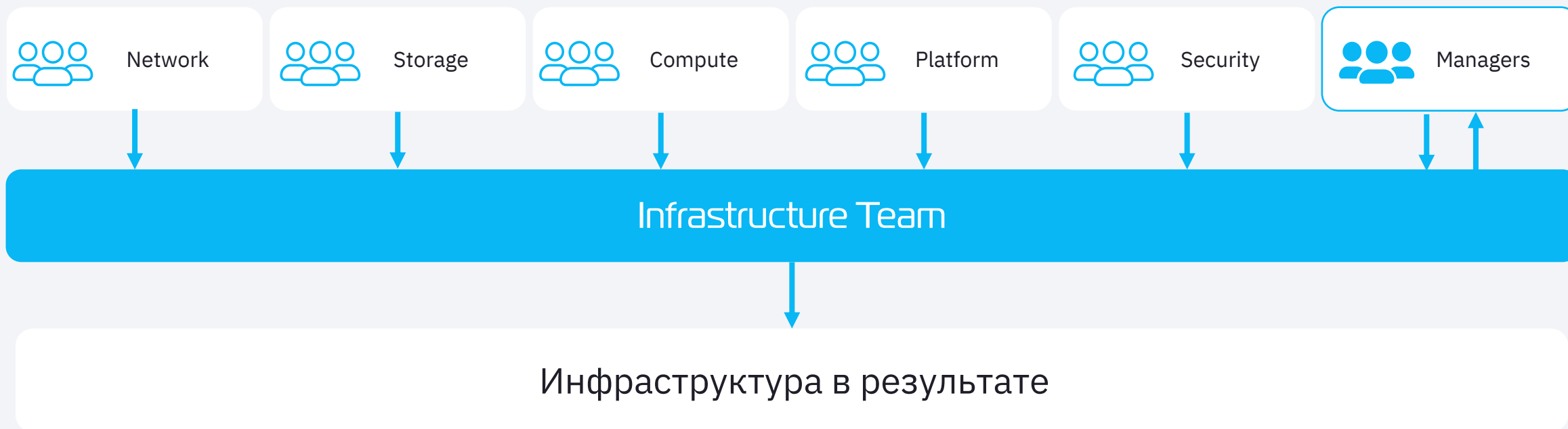
Ответственные за инфраструктуру



«Единое окно» по вопросам инфраструктуры для других команд

- Решение оперативных задач по инфраструктуре
- Планирование и сопровождение работ в ЦОД
- Установка и ввод оборудования для работы сервисов Облака
(либо его демонтаж и вывод)
- Тестирование нового оборудования
- Hardware-мониторинг

Новый процесс взаимодействия команд



К чему хотим прийти в инфраструктуре

К чему хотим прийти в инфраструктуре



Унифицировать подключения серверов и стоек

К чему хотим прийти в инфраструктуре

- ✓ Унифицировать подключения серверов и стоек
- ✓ Минимизировать кол-во линков выходящих из стойки

К чему хотим прийти в инфраструктуре

- ✓ Унифицировать подключения серверов и стоек
- ✓ Минимизировать кол-во линков выходящих из стойки
- ✓ Развернуть все коммутаторы портами назад

К чему хотим прийти в инфраструктуре

- ✓ Унифицировать подключения серверов и стоек
- ✓ Минимизировать кол-во линков выходящих из стойки
- ✓ Развернуть все коммутаторы портами назад
- ✓ Распределить оборудование по стойке с учетом энергопотребления

К чему хотим прийти в инфраструктуре

- ✓ Унифицировать подключения серверов и стоек
- ✓ Минимизировать кол-во линков выходящих из стойки
- ✓ Развернуть все коммутаторы портами назад
- ✓ Распределить оборудование по стойке с учетом энергопотребления
- ✓ Минимизировать пустое пространство

К чему хотят прийти в смежных командах

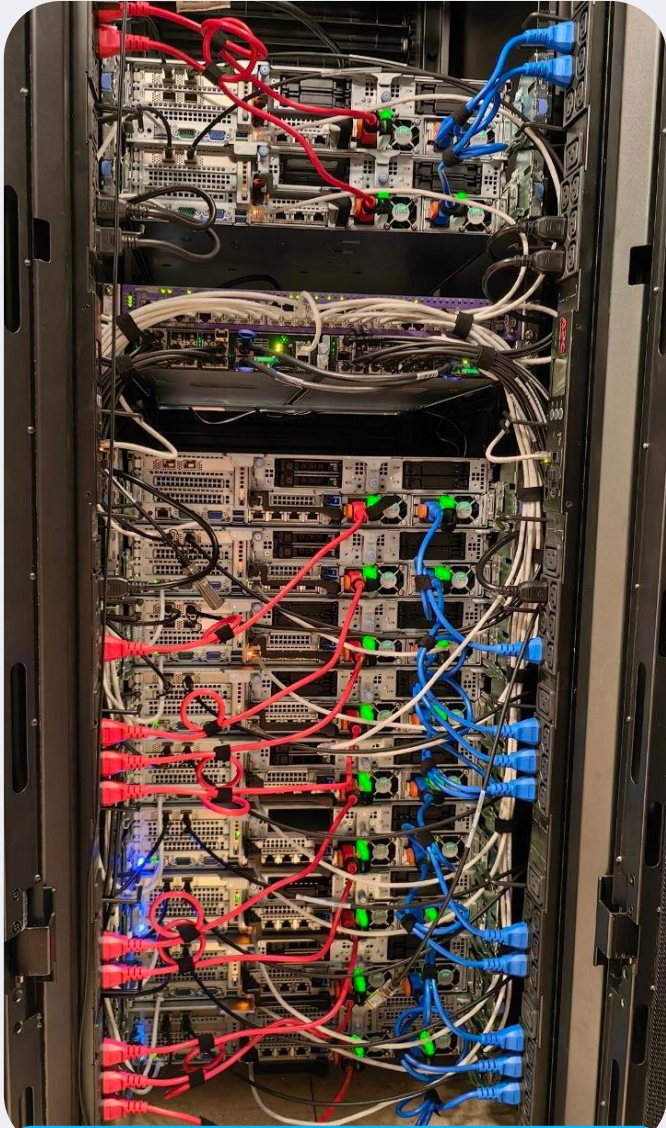
- Организовать отдельные сетевые стойки
- Отказаться от InfiniBand в пользу Ethernet, изменить архитектуру сети
- Отказаться от СХД в пользу распределённых систем хранения



Стандарты построения инфраструктуры



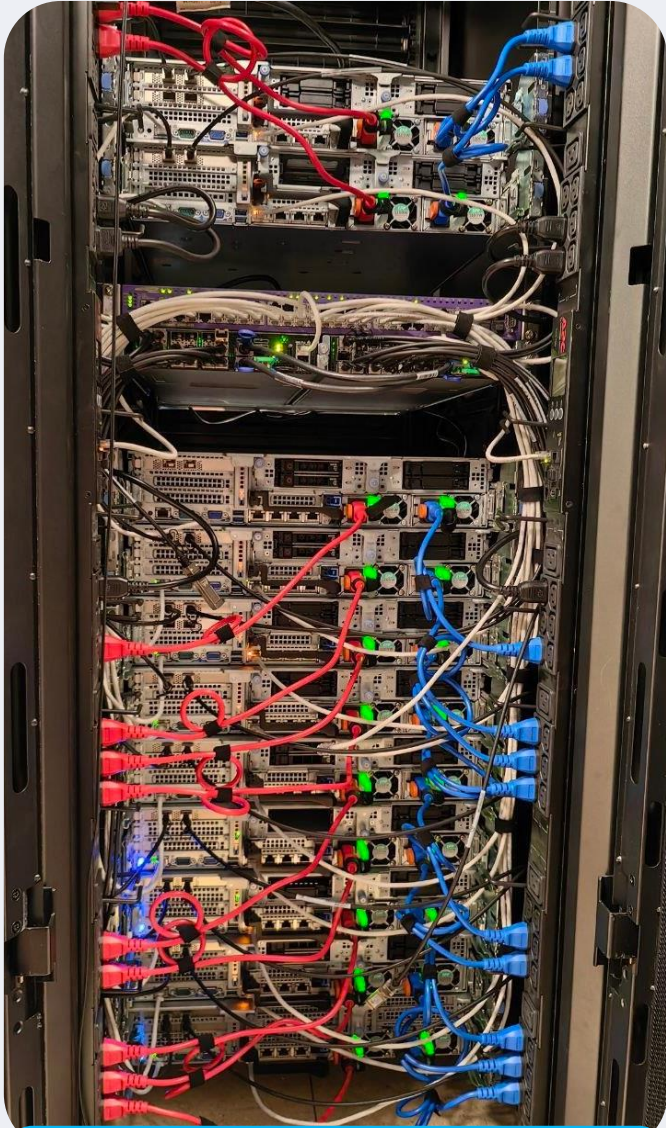
Классификация стоек



Cloud racks

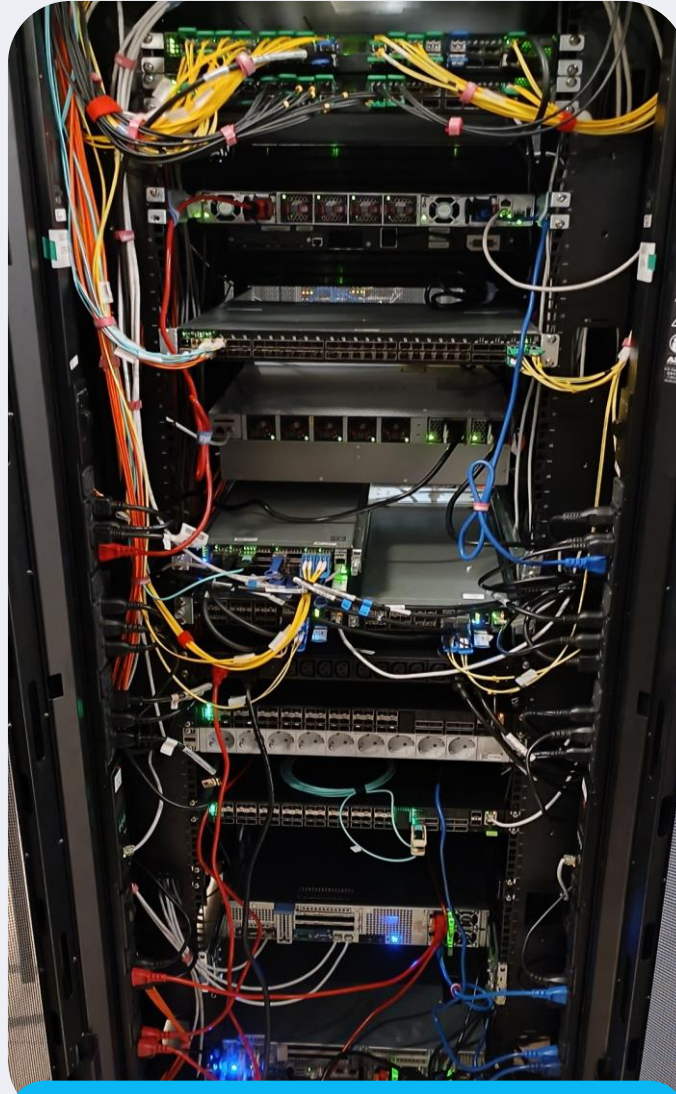
Здесь работает Облако

Классификация стоек



Cloud racks

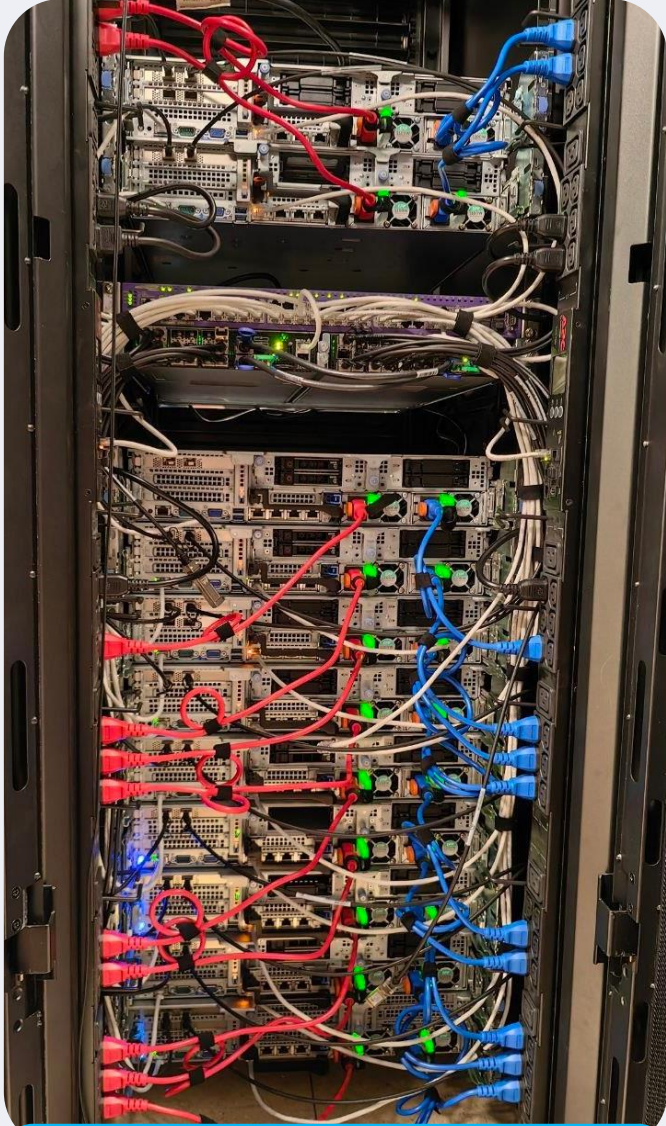
Здесь работает Облако



Network racks

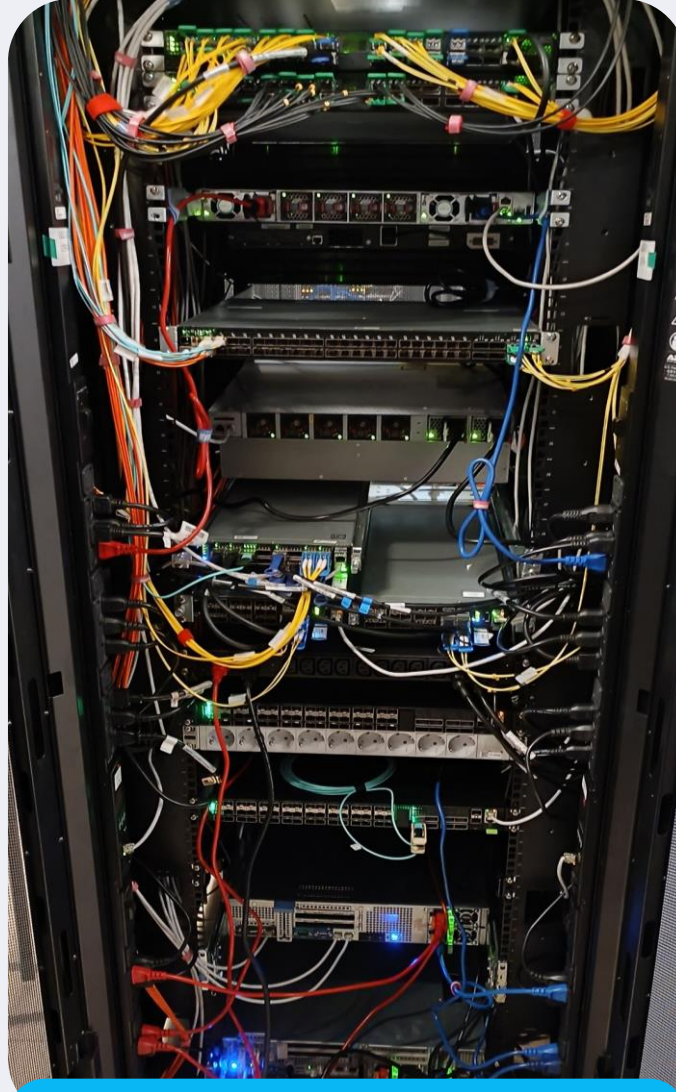
Здесь находится ядро сети и все
внешние подключения к Облаку

Классификация стоек



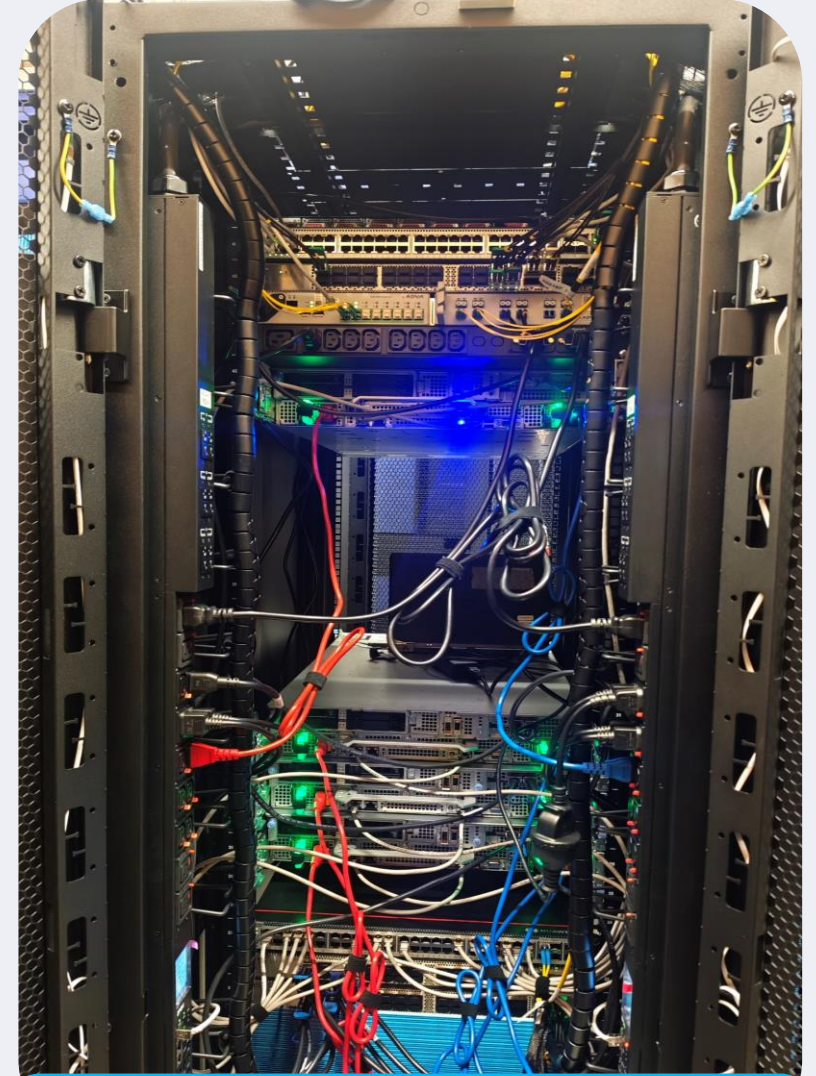
Cloud racks

Здесь работает Облако



Network racks

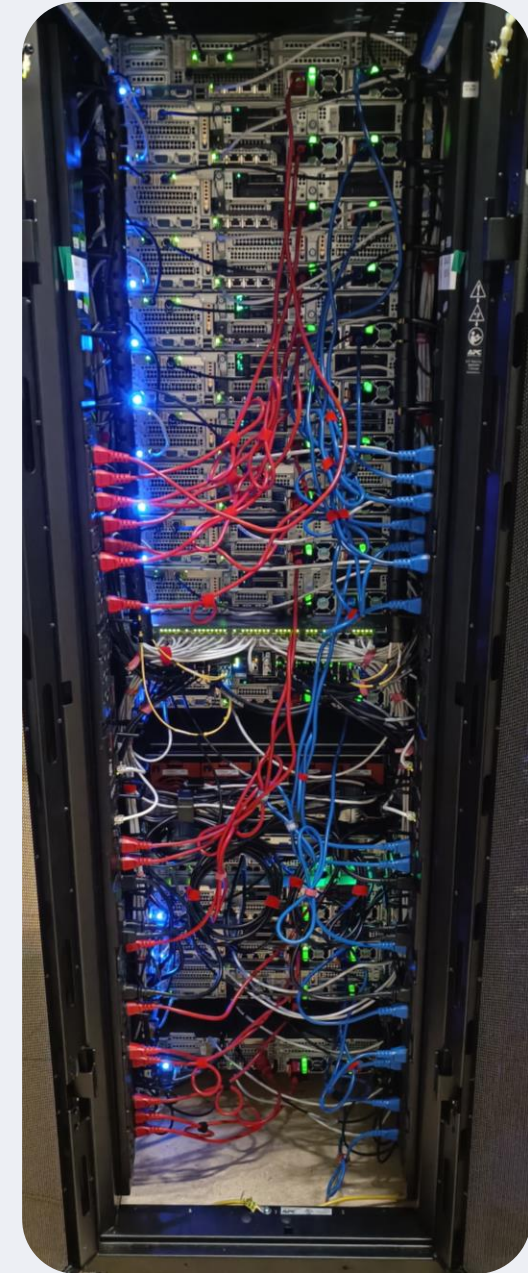
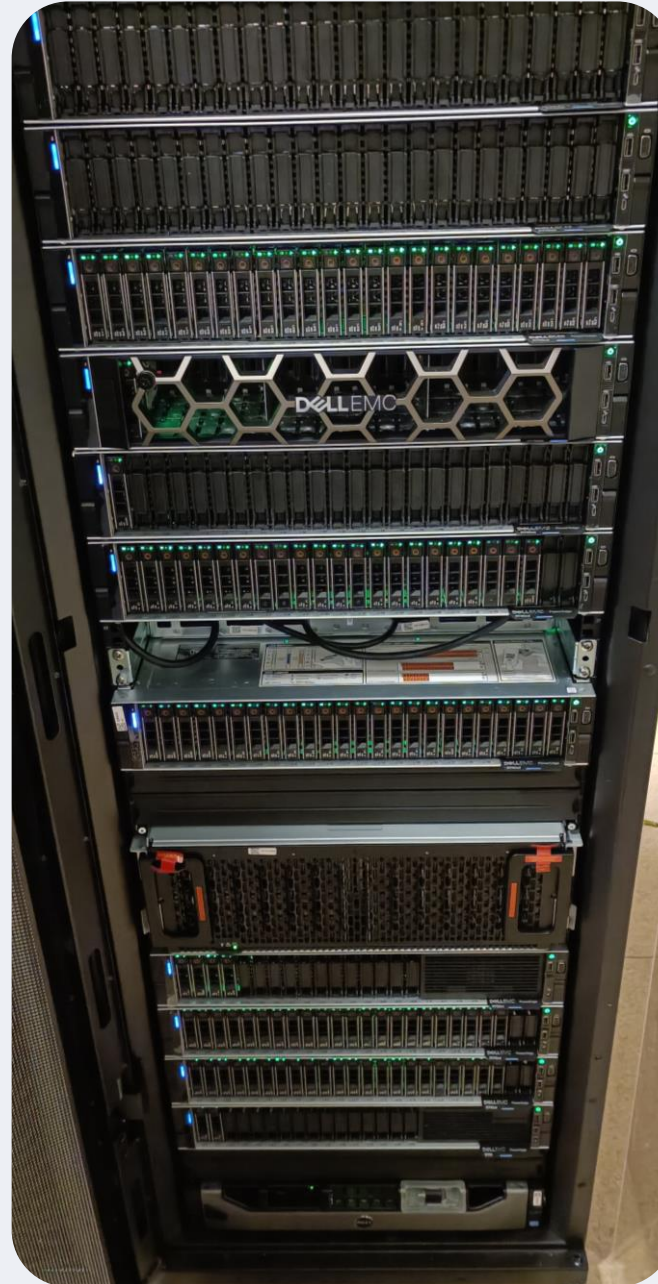
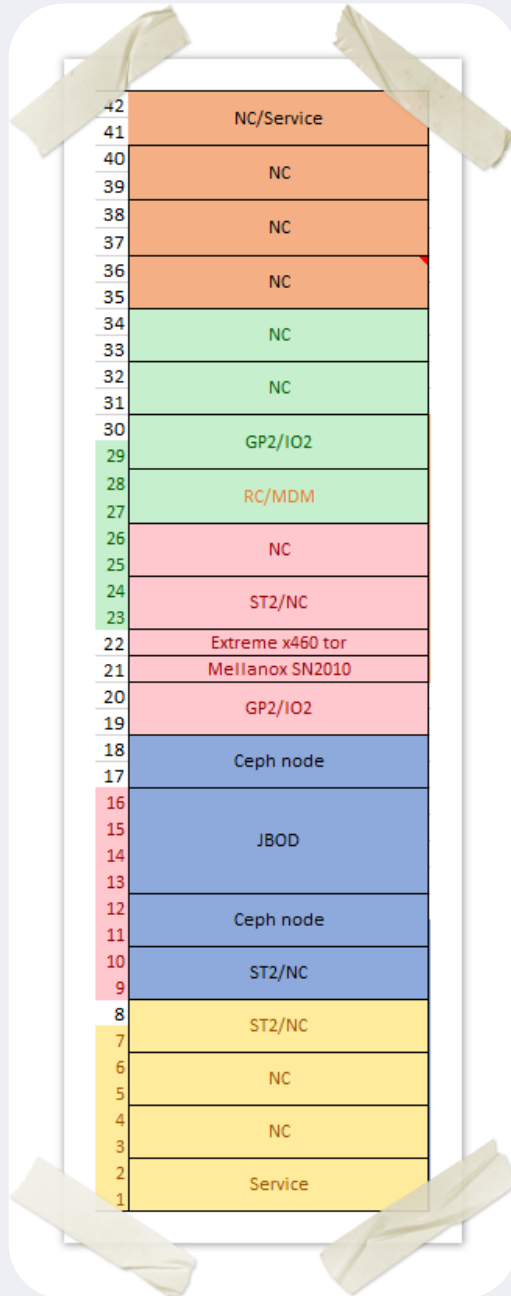
Здесь находится ядро сети и все внешние подключения к Облаку



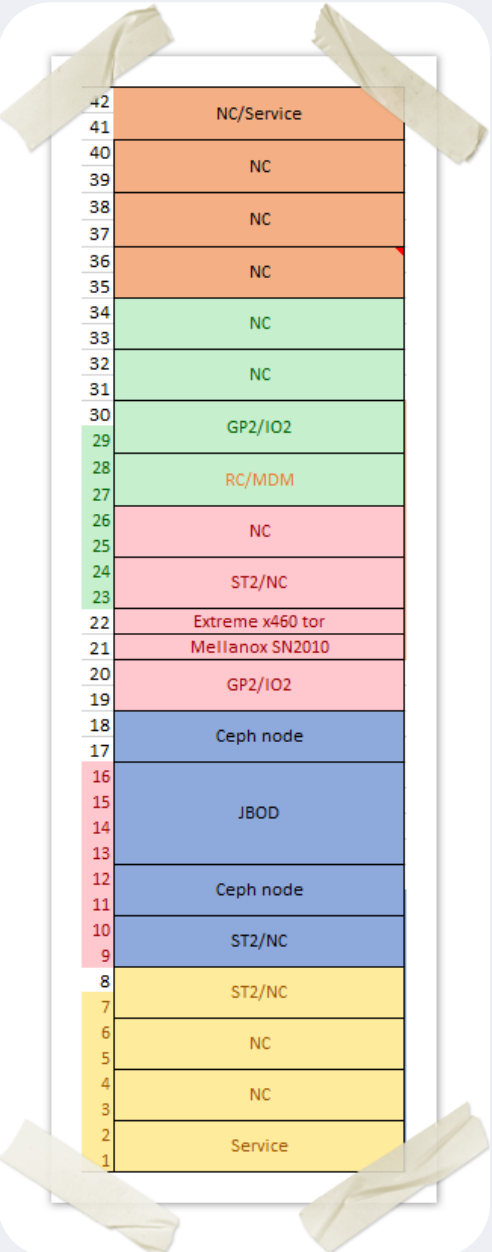
Combined racks

Совмещение первых двух вариантов. Используются в небольших инсталляциях

Единый стандарт сборки стоек



Распределение питания. Power Distribution Units



Стойка	PDU 1, A			PDU 2, A	
	A	B	C	D	E
	14	15,8	13,5	15,4	14,8

Стандарт на основе расчетов

Спецификация питания

Электрический щитЩБ-А1.7

Кол-во фаз1p3p

Номер автоматаQF30

Сила тока16 A

Фактическая мощность (Вт)Фактическая нагрузка (A)

На фазу 1 3,6 A / 16 A

23%

На фазу 2: 4,4 A / 16 A

28%

На фазу 3: 4,8 A / 16 A

30%

Напряжение:229 B

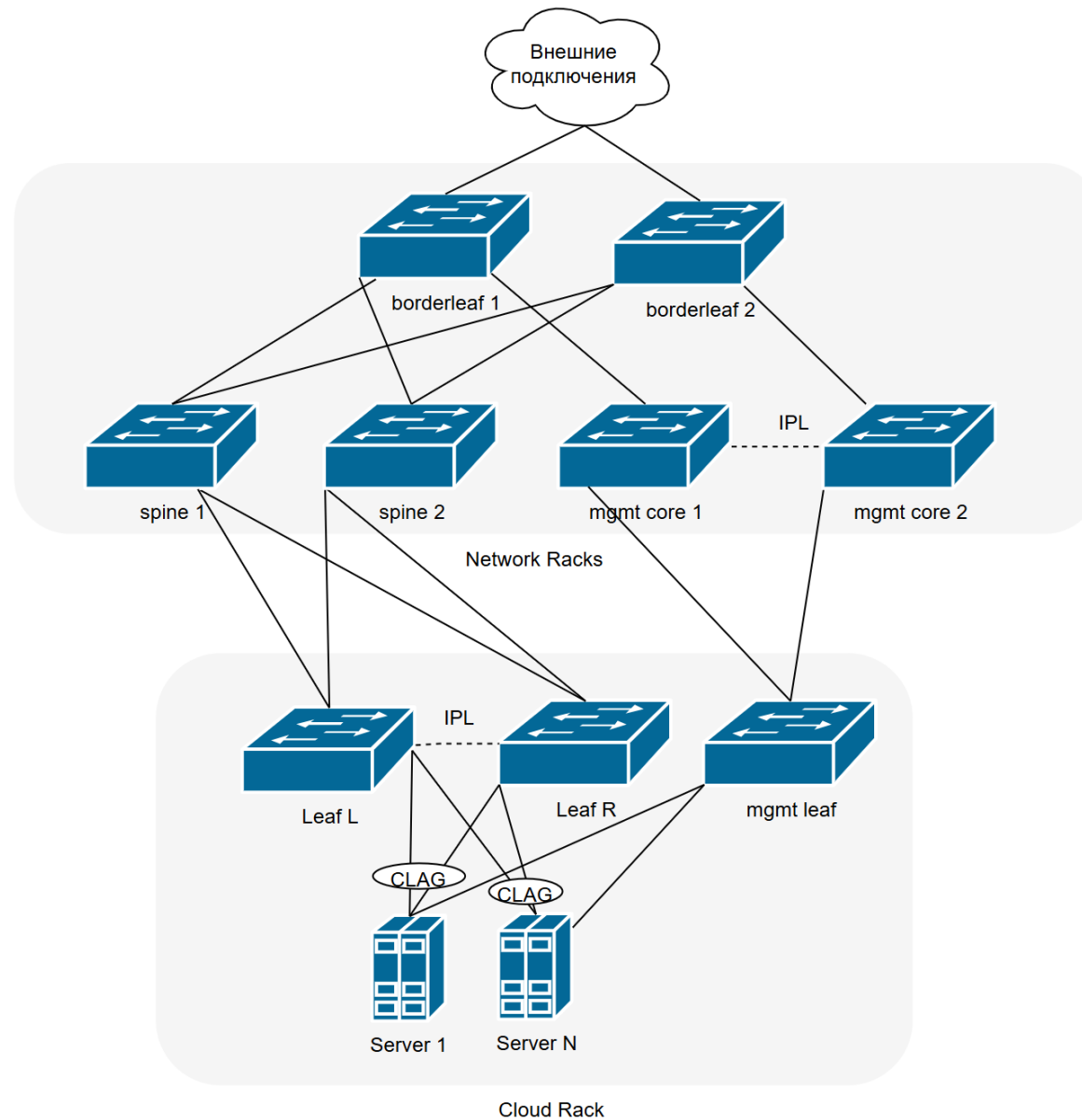
Время работы:192 days, 21 hours, 32 minutes and 49 seconds

Версия прошивки:v6.9.6

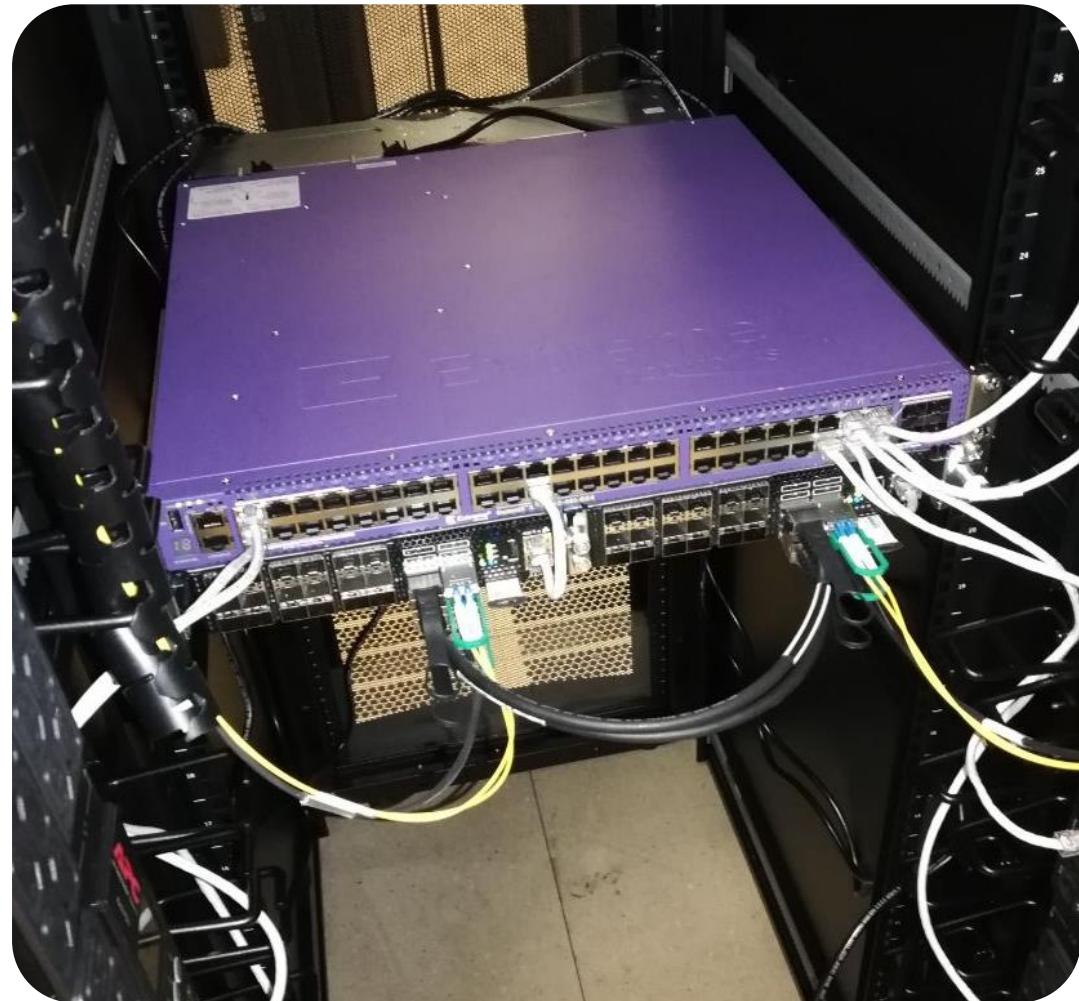
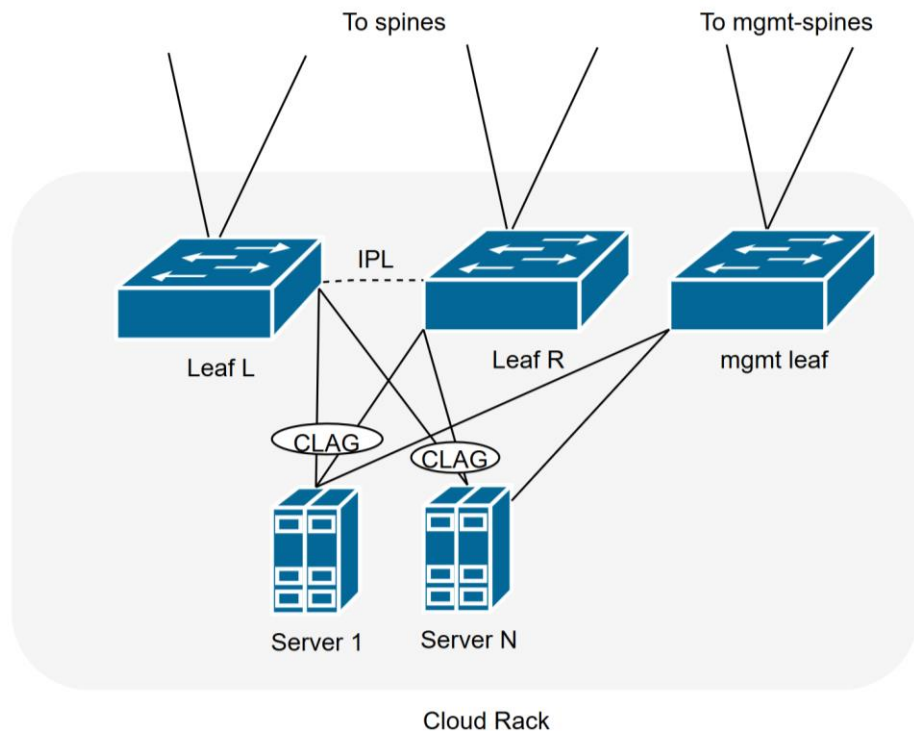
Мониторинг (DCIM + Zabbix)

38

Новая сеть

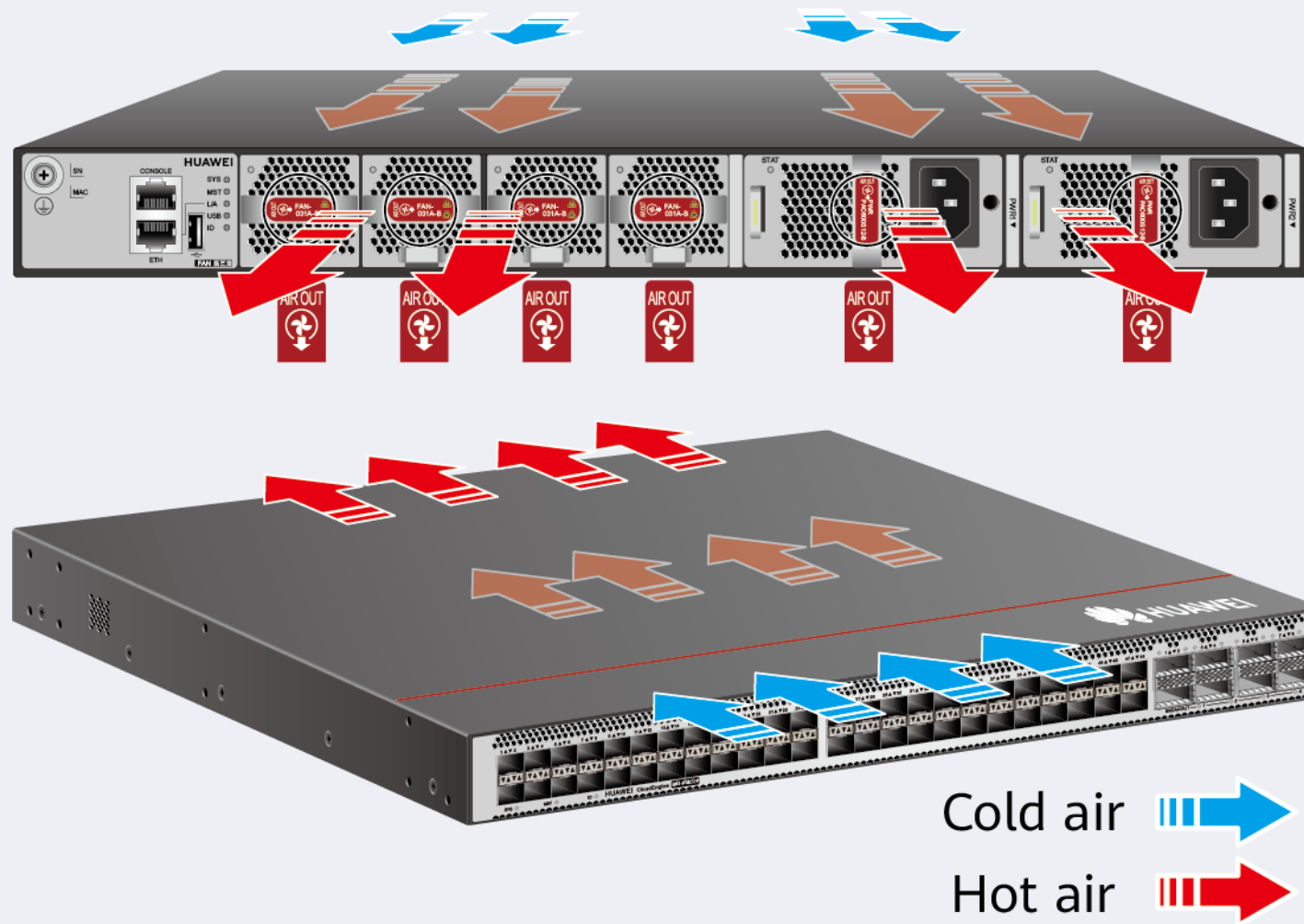


Коммутация на старте Cloud Rack

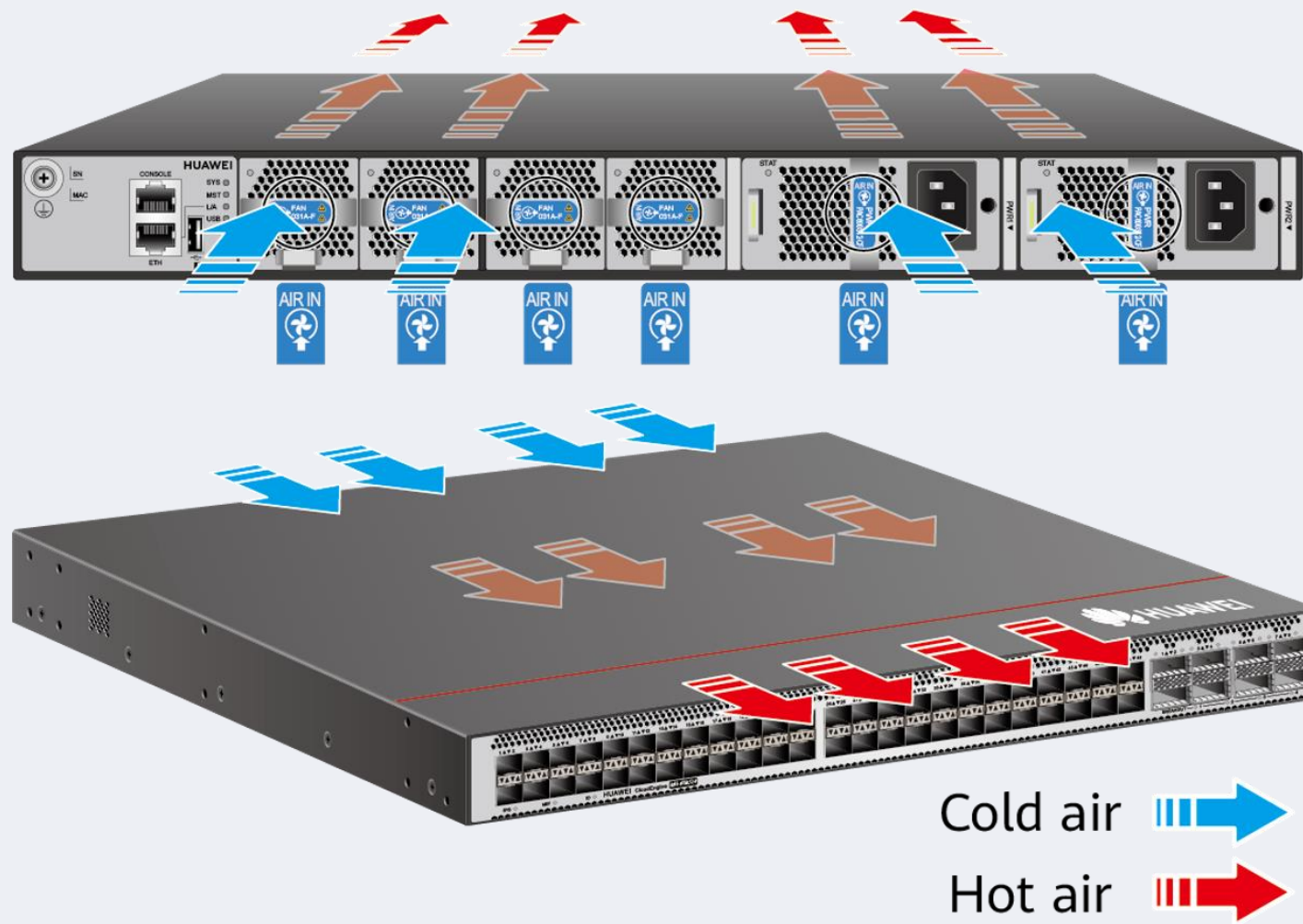


6 линков выходит из стойки

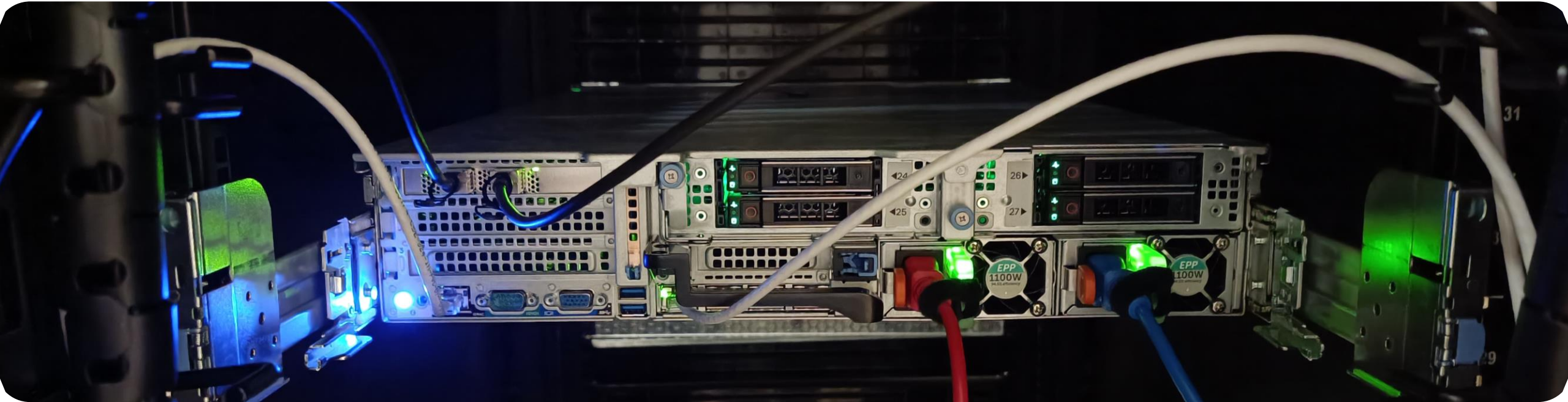
Обдув коммутатора при установке спереди



Обдув коммутатора при установке сзади



Коммутация любого сервера внутри Cloud Rack



25G/100G
(Mellanox CX5/CX6)

eth4 -> leaf-L
eth5 -> leaf-R

Cluster (untagged)
+ additional vlans
(tagged)



1G

BMC -> mgmt-leaf - Manage
eth0 -> mgmt-leaf - ClusterBackup

Фрагмент плана коммутации

План коммутации по стойке

Роль	Юнит	Порт						
		idrac mgmt	clusterbackup	cluster	SAS to JBOD	L3 с Mellanox SN2700	peerlink	intercluster
NC/Service	41-42	21	23	1	-	-	-	-
NC	39-40	19	25	4	-	-	-	-
NC	37-38	17	27	7	-	-	-	-
NC	35-36	15	29	10	-	-	-	-
NC	33-34	13	31	13	-	-	-	-
NC	31-32	11	33	16	-	-	-	-
GP2	29-30	9	35	2	-	-	-	-
RC/MDM	27-28	7	37	5	-	-	-	-
NC	25-26	5	39	8	-	-	-	-
NC/SDS(st2)	23-24	3	40	11	-	-	-	-
Extreme x460	22	-	-	-	-	-	-	47, 48
Mellanox SN2010	21 R	44	-	-	-	21, 22	19, 20	-
Mellanox SN2010	21 L	22	-	-	-	21, 22	19, 20	-
SDS(io2)	19-20	4	38	14	-	-	-	-
Ceph node / Service	17-18	6	36	17	A4	-	-	-

Конфигурация в hiera



```
5517 se190
5518   cobbler_profile:
5519   zone: rack-af03
5520   mgmt:
5521     type: idrac
5522     mac: '70:b5:
5523     ip: 10.
5524   ifaces:
5525     clusterbackup:
5526       network: clusterbackup
5527       ip: 10.
5528       slaves: [eth0]
5529       route_table: 200
5530     cluster:
5531       network: cluster
5532       mac: ['08:c0:', '08:c0:']
5533       ip: 10.
5534       slaves: [eth4, eth5]
5535       bond_mode: 802.3ad xmit_hash_policy=layer3+4 miimon=100
5536     storage:
5537       vid: 12
5538       physdev: cluster
5539       network: storage
5540       ip: 10.
```

Описание интерфейсов сервера

```
Y cumulus-leaf-af03-l. .yaml X
11   c2_cumulus::interfaces:
12   bond:
89   se190:
90     mtu: 9216
91     clag_id: 11
92     pvid: 10
93     vids: [12]
94     slaves: [swp11]
95     lacp_bypass_allow: 1
```

Описание порта на свиче (Cumulus)

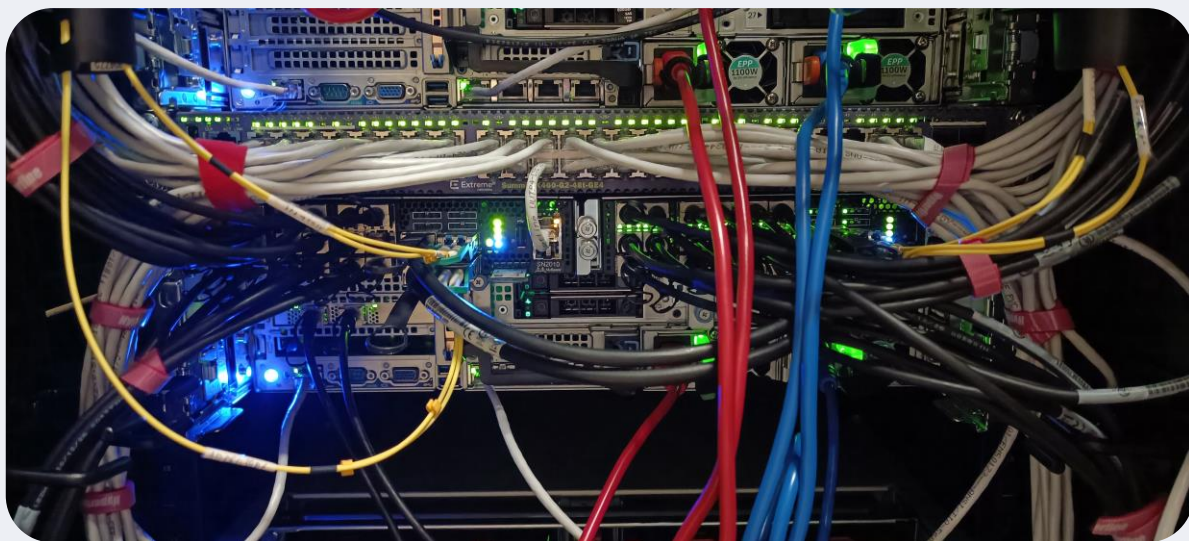
Конфигурация в



```
mgmtleaf-xlrt1-52-k-09. [redacted] yml x
1  ---
2  networks:
3    vlans:
4      11: manage
5      14: clusterbackup
6    vrfs:
7      _management_vpn_:
8        - family: ipv4-family
9    interfaces:
10     LoopBack0:
11       ipv4: [redacted]
12       mgmt: true
13     Vlanif11:
14       description: "<<< MANAGE network >>>"
15       ipv4: [redacted]
16     Vlanif14:
17       description: "<<< CLUSTERBACKUP network >>>"
18       ipv4: [redacted]
19     GE1/0/2:
20       description: "PDU(3f) L"
21       pvid: 11
22     GE1/0/5:
23       description: "sx012m"
24       pvid: 11
```

Описание портов на свиче (Проприетарные ОС)

Завершенная коммутация



Cloud Rack (Leafs)



Network Rack (Spines)

Ну а теперь про автоматизацию

(ну наконец-то)

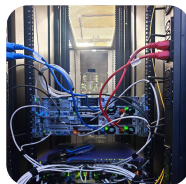


Этапы ввода нового сервера в Облако

Этапы ввода нового сервера в Облако

1

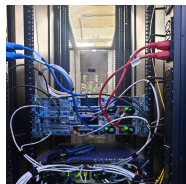
Определение
места установки



Этапы ввода нового сервера в Облако

1

Определение
места установки



2

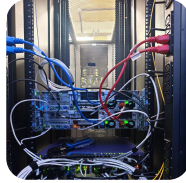
Подготовка и выкатка
конфигов



Этапы ввода нового сервера в Облако

1

Определение
места установки



2

Подготовка и выкатка
конфигов



3

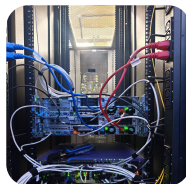
Подготовка сервера



Этапы ввода нового сервера в Облако

1

Определение
места установки



4

Разливка ОС



2

Подготовка и выкатка
конфигов



3

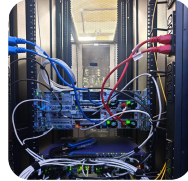
Подготовка сервера



Этапы ввода нового сервера в Облако

1

Определение
места установки



4

Разливка ОС



2

Подготовка и выкатка
конфигов



5

Ввод в облако



3

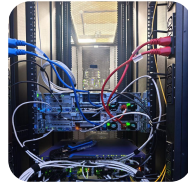
Подготовка сервера



Этапы ввода нового сервера в Облако

1

Определение
места установки



4

Разливка ОС



2

Подготовка и выкатка
конфигов



5

Ввод в облако



3

Подготовка сервера



6

Проверка



1. Определение места установки



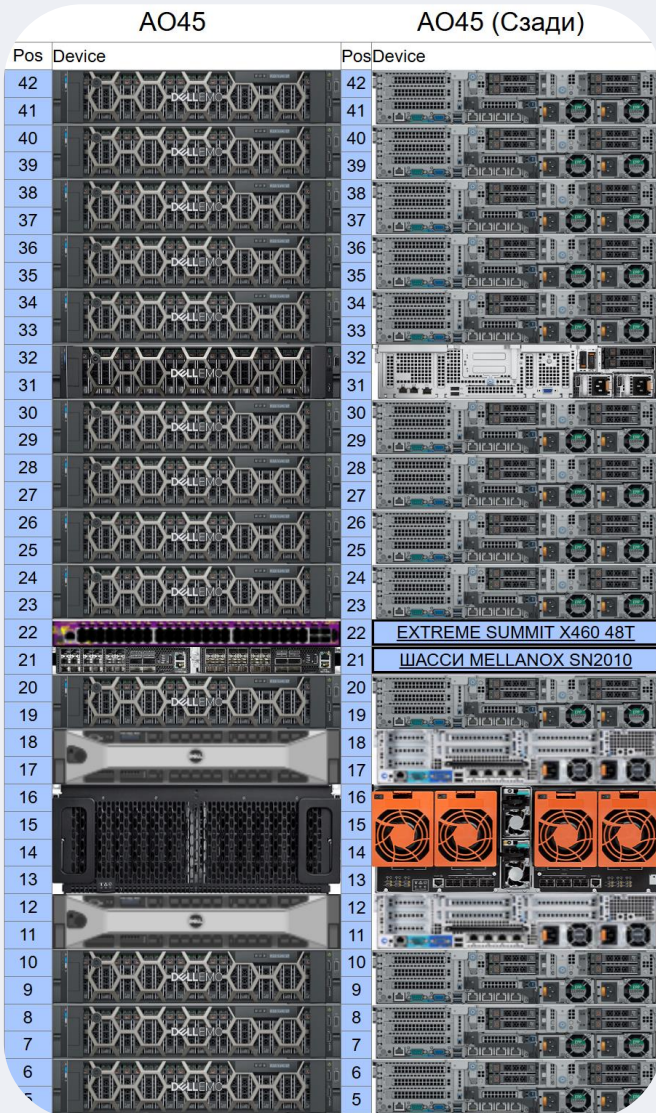
Выбор свободного места —
в соответствии со стандартом



Физические работы в ЦОД

1. Определение места установки

Система инвентаризации DCIM и работа с её API



GET

/device Information about one or more devices

POST

/device/{DeviceID} Update the information for a specific device

POST

/device/{DeviceID}/store Move the specified device to the storage room

```
"error": false,
"errorcode": 200,
"device": {
  "DeviceID": 17721,
  "Label": "DELL POWEREDGE R740XD",
  "SerialNo": "GXD1JH3",
  "Owner": 5,
  "Cabinet": 2101,
  "Position": 33,
  "Height": 2,
  "TemplateID": 42,
  "CLOUD_hostname": "se365",
  "Controller_model": "PERC H730P Mini",
  "CPU_model": "Intel(R) Xeon(R) Gold 6254 CPU @ 3.10GHz",
  "CPU_number": "2",
  "GPU_kit": "off",
  "GPU_model": "",
```

1. Определение места установки

Разработанный консольный клиент OpenDCIM – облегчает определение местоположения

```
[19:16]:[ ~]# opendcim show rack -n af45
```

NAME	DEVICE	SERIAL	UNIT	ROLES
OWNER: [REDACTED] DATACENTER: Компрессор [REDACTED], RACK: AF45				
sf240	DELL POWEREDGE 740XD	GV53YW3	41-42	nc (gen5hd)
sf166	ДАТАРУ ПИ760	4V8F954	39-40	nc (gen7-dh), dedicated
sf165	ДАТАРУ ПИ760	48T5954	37-38	nc (gen7-dh), dedicated
sf164	ДАТАРУ ПИ760	407F954	35-36	nc (gen7-dh), dedicated
sf163	ДАТАРУ ПИ760	3V8F954	33-34	nc (gen7)
sf239	DELL POWEREDGE 740XD	HZTT6X3	31-32	nc (gen5hd)
sf010	ДАТАРУ ПИ760	G7KGX64	29-30	ceph.mon.rbd-ru-msk-comp1p-01, ceph.osd.rbd-ru-msk-comp1p-01
sf120	DELL POWEREDGE R720	3RTL5Y1	27-28	ovn.l3gw
sf162	ДАТАРУ ПИ760	3P1NH54	25-26	nc (gen7-dh), dedicated
	FREE (2)		23-24	
	ДАТАРУ C-X460-2	1926G-00832	22	
	MELLANOX SN2010 LEFT	MT2149J14441	21/1	
	MELLANOX SN2010 RIGHT	MT2413T001C8	21/2	
sf104	ДАТАРУ ПИ760	CSBBJB4	19-20	sio_sds
	FREE (4)		15-18	

```
[19:17]:[ ~]# opendcim show device -n sf020 -w cpu ram
```

NAME	DEVICE	SERIAL	DATACENTER	RACK	UNIT	ROLES	CPU	RAM
sf020	ДАТАРУ ПИ760	HN1NH54	Компрессор-[REDACTED]	AF45	3-4	nc (gen7gl50)	2 x INTEL(R) XEON(R) GOLD 6548Y+	2048 Gb (32x64 Gb)

```
[19:17]:[ ~]#
```

1. Определение места установки

Утилита для поиска свободных юнитов с нужной ролью по всей AZ

```
10:45]:[]# c2-prepare-host units -r service
Processing 100g racks...
Processing ██████████

Free units in ██████████ an18 :
- 3-4 (service)
- 1-2 (service/efs/baas/nvme)
Processing 25g racks...

Free units in ak23:
- 41-42 (nc/service)

Free units in an23:
- 41-42 (nc/service)

Free units in an25:
- 41-42 (nc/service)

Free units in ay20:
- 41-42 (nc/service)

Free units in an26:
- 41-42 (nc/service)

Free units in ay22:
- 41-42 (nc/service)
```

Время поиска сократилось с 30 минут до 30 секунд

1. Определение места установки

Модуль puppet для отправки инвентарной информации

```
exec { 'inventory_collector.py':  
  command    => '/bin/python /usr/local/puppet-bin/inventory_collector.py -f /run/inventory_collector/inventory.json.tmp',  
  logoutput  => true,  
  timeout    => 10,  
  user       => 'inventory-collector',  
  group      => 'inventory-collector',  
  creates    => '/run/inventory_collector/inventory.json.tmp',  
}  
~> exec { 'inventory_pusher.py':  
  command    => "/bin/python /usr/local/puppet-bin/inventory_pusher.py -f /run/inventory_collector/inventory.json.tmp -t",  
  logoutput  => true,  
  timeout    => 10,  
  user       => 'inventory-collector',  
  group      => 'inventory-collector',  
  refreshonly => true,  
}
```

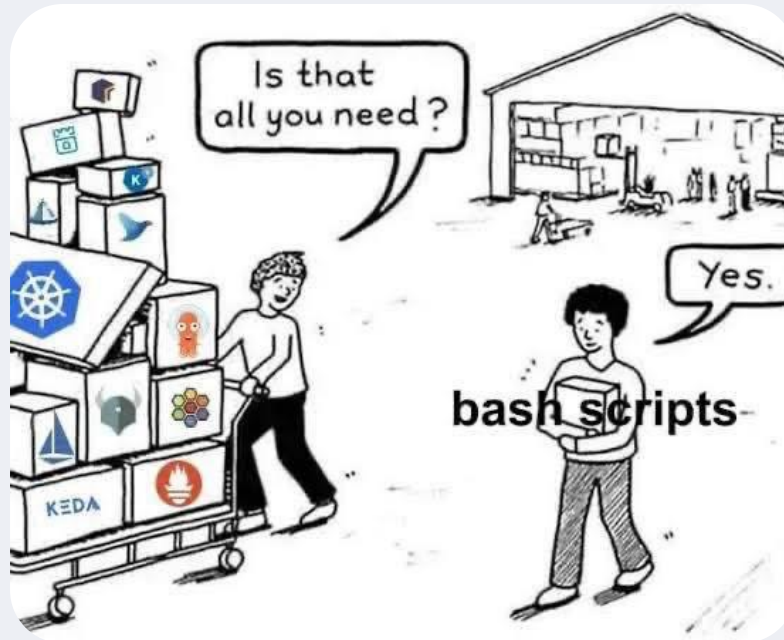


```
"CLOUD_hostname": "sg100",  
"CPU_model": "Intel(R) Xeon(R) Gold 6254 CPU @ 3.10GHz",  
"CPU_number": "2",  
"Controller_model": "PERC H750 Adapter",  
"GPU_model": "",  
"GPU_number": "0",  
"Installed_BOSS": "off",  
"Installed_Disks": "2 disks ( 2x558 Gb )",  
"Installed_RAM": "512 Gb ( 8x64 Gb )",  
"Model": "PowerEdge R740",  
"Serial_number": "682T6L3"
```

2. Подготовка конфигов



2. Подготовка конфигов



2. Подготовка конфигов



```
avtodns x
C:\Users\anton\AppData\Local\Programs\Python\Python312\python.exe C:
(づ｡◕‿｡◕｡)づ removing mode is enabled - check results before commit!
processing host sf600
Host sf600 skipped. Some addresses is already used

processing host sf601
ADDED: sf601 IN A 10.
ADDED: sf601s IN A 10.
ADDED: sf601m IN A 10.
ADDED: sf601b IN A 10.

processing host sf602
```

```
Terminal - repo-c2ansible
PS C:\> python auto_ansible.py remove -n sg269 -z
-r ay25 -u 27-28 -p ./ansible
ansible\inventories\ \host_vars\mgmt-leaf-ay25\
Lines to remove:
7:
  pvid: 11
  alias_name: sg269m
37:
  pvid: 14
  alias_name: sg269b
Do you want to continue? [y/N]: y
Done!
PS C:\> python auto_ansible.py add -n sg600 -z
ay26 -u 27-28 -p ./ansible
ansible\inventories\ \host_vars\mgmt-leaf-ay26\
Clusterbackup port: 37
iDRAC port: 7
Done!
PS C:\> repo-c2ansible>
```

```
> python3 ./auto_hiera.py add -cif -ip -P ../hiera-k2 -hp 732000
Enter roles. One role per line! Enter an empty string when all desired roles are added. Available roles are ['nc', 'sio_sds']
Enter role #1: nc
Enter role #2:
Enter hostname: se888
AZ is
Enter domain without hostname and az. Default is [ ]:
Enter rack: ai08
Enter cobbler profile:
Enter mac of first cluster slave interface: 11:22:33:44:55:66
Enter mgmt interface mac: aa:bb:cc:dd:ee:ff
Enter 2 cluster slaves. Default is ['eth2', 'eth3']:
Enter unit: 33-34
Default vlans for ['nc'] are [12]
Do you want to use defaults? [Y/n]
hiera-k2\hieradata\ \cumulus-leaf-ai08-1. .yaml
Cluster port: 13
hiera-k2\hieradata\ \cumulus-leaf-ai08-n. .yaml
Cluster port: 13
Do you want to write cumulus files? [y/N]: y
Left is Done!
Right is Done!
se888
cobbler_profile: centos-8.4-x86_64
zone: rack-ai08
mgmt:
  type: idrac
  mac: 'aa:bb:cc:dd:ee:ff'
```

Значительно снижается риск ошибок «человеческого фактора»

3. Подготовка сервера



Обновление прошивок

Настройка BIOS, BMC

Настройка Storage

PXE Boot

3. Подготовка сервера

Пример веб-интерфейса BMC контроллера (Dell iDRAC)

The screenshot displays the Dell iDRAC9 Enterprise web interface. The top navigation bar includes links for Dashboard, System, Storage, Configuration, Maintenance, and iDRAC Settings. A search bar and user profile icon are on the right. Below the navigation bar, there are action buttons for Graceful Shutdown, LED On, and More Actions. The main content area is divided into three sections: Health Information, System Information, and Task Summary.

Health Information

SYSTEM IS HEALTHY

System Health: Healthy (Details)

Storage Health: Healthy (Details)

System Information

Power State	ON
Model	PowerEdge R740xd
Host Name	
Operating System	
Operating System Version	
Service Tag	
BIOS Version	2.9.4
iDRAC Firmware Version	4.40.00.00
IP Address(es)	10. [redacted]
iDRAC MAC Address	f4:ee: [redacted]
License	Enterprise Edit

Task Summary View All Jobs

Pending Jobs : 0
No Pending Jobs

In-Progress Jobs : 0
No In-Progress Jobs

Completed Jobs : 9
0 with Errors
1 Failed

Recent Logs view all

Severity	Description	Date and Time
✓	Log cleared.	Mon 07 Aug 2023 11:50:49

Virtual Console Settings

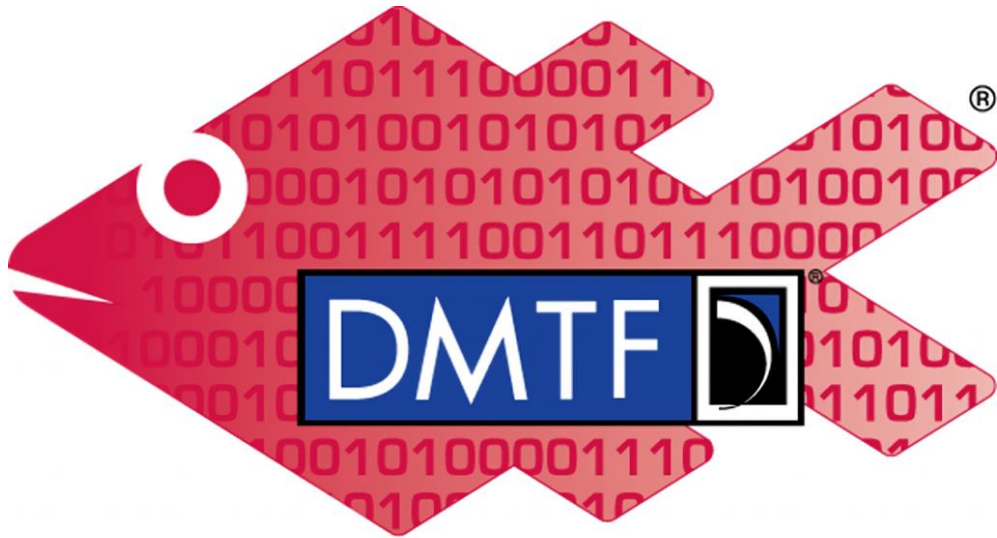
```

iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB
iDRAC9 (10.10.10.10) ScaleIO iDRAC9-Create-1100 - Created device script (10.10.10.10) - Capacity 67888MB LB

```

3. Подготовка сервера

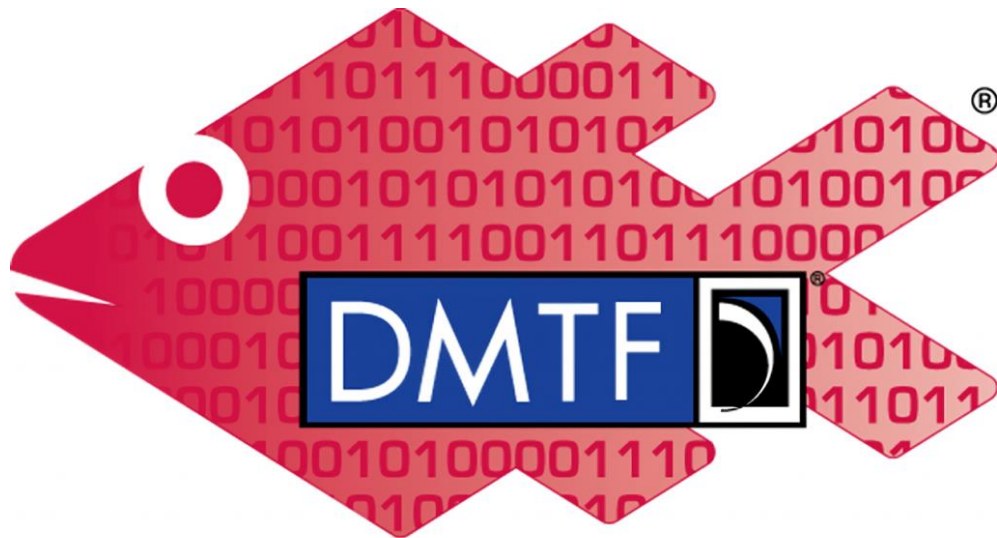
Redfish API – универсальный механизм взаимодействия с BMC



Redfish

3. Подготовка сервера

Redfish API – универсальный механизм взаимодействия с BMC



Redfish



Подходит в полной мере

- ASUS
- Huawei
- Dell
- YADRO
- HPe
- MadTech



Подходит с нюансами

- GOOXI
- SuperMicro

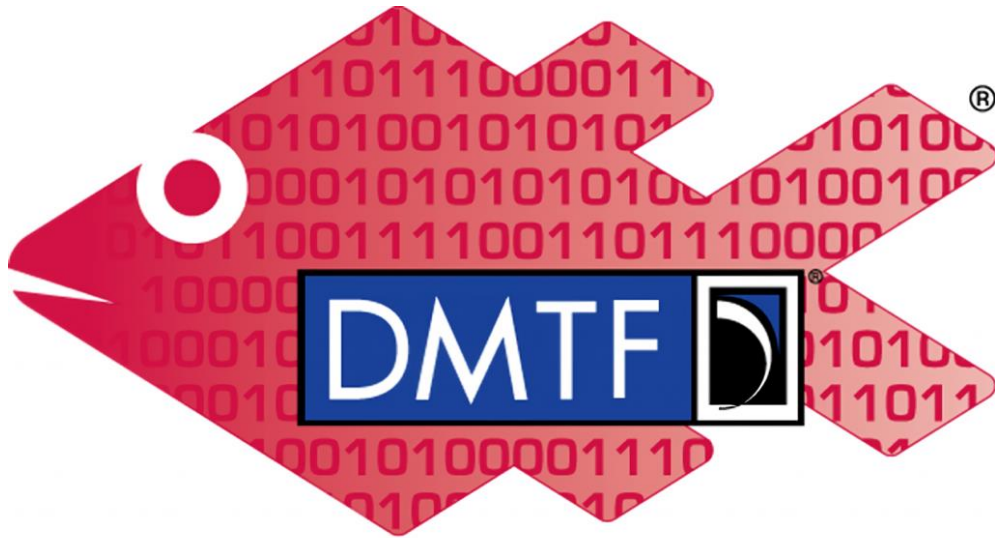


Не подходит

- Aquarius
- DEPO

3. Подготовка сервера

Redfish API – пример запросов для выполнения PXE boot в OpenBMC



Redfish

```
curl -k -u логин:пароль -H "Content-Type: application/json" -  
X PATCH https://server.test/redfish/v1/Systems/system -d  
{  
  "Boot": {  
    "BootSourceOverrideEnabled": "Once",  
    "BootSourceOverrideTarget": "Pxe"  
  }  
}
```

```
curl -k -u логин:пароль -H "Content-Type: application/json" -  
X POST https://server.test/redfish/v1/Systems/system/  
Actions/ComputerSystem.Reset -d  
{  
  "ResetType": "GracefulRestart"  
}
```

3. Подготовка сервера

Разработанная утилита c2-prepare-host:

- Обновляет прошивки BMC, BIOS
- Устанавливает режим загрузки UEFI/Legacy
- Настраивает параметры BIOS и BMC с учетом будущей роли сервера
- Собирает RAID-1 под ОС (если нужно)
- Отправляет сервер в PXE boot

• Время подготовки сервера сокращается с 1 часа до 15 мин

• ... или больше если готовим одновременно несколько штук

```
=====
Processing host 172.29.34.10
=====
Enter IDRAC current password:
Serial number is B82T6L3, please verify it's correct (y/n): y
2024-04-11 18:31:43,583 Checks successfully done

Detected server model: R740, Detected server state: Off
Config for node loaded successfully

=== [172.29.34.10] Step 1 of 7: Update IDRAC password ===

Enter new IDRAC password (leave blank if you don't need to change it):
SKIPPED: IDRAC password change skipped

=== [172.29.34.10] Step 2 of 7: Update IDRAC version ===

Starting update IDRAC from version 4.40 to version 6.10
It will take about 5-10 minutes, wait...
2024-04-11 18:32:46,125 POST method returned 201 status code, all is fine / UpdateService/FirmwareInventory
2024-04-11 18:32:46,125 Available-25227-6.10.80.00 __iDRAC.Embedded.1-1 was successfully uploaded to iDRAC
2024-04-11 18:32:47,713 POST method returned 202 status code, all is fine / UpdateService/Actions/Update
2024-04-11 18:38:17,692 Success
Update done successfully, current version is 6.10.80.00

=== [172.29.34.10] Step 3 of 7: Update BIOS version ===

Starting update BIOS from version 2.9.4 to version 2.17.1
It will take about 5-10 minutes, wait...
2024-04-11 18:38:38,135 POST method returned 201 status code, all is fine / UpdateService/FirmwareInventory
2024-04-11 18:38:38,136 Available-159-2.17.1 __BIOS.Setup.1-1 was successfully uploaded to iDRAC
2024-04-11 18:38:40,699 POST method returned 202 status code, all is fine / UpdateService/Actions/Update
2024-04-11 18:38:43,393 POST method returned 204 status code, all is fine / Systems/System.Embedded.1/Action
2024-04-11 18:43:43,832 Current job state is not done, will retry in 1 min (attempt 1/15)
2024-04-11 18:44:44,052 Current job state is not done, will retry in 1 min (attempt 2/15)
2024-04-11 18:45:44,323 Success

=== [172.29.34.10] Step 4 of 7: Set IDRAC parameters ===

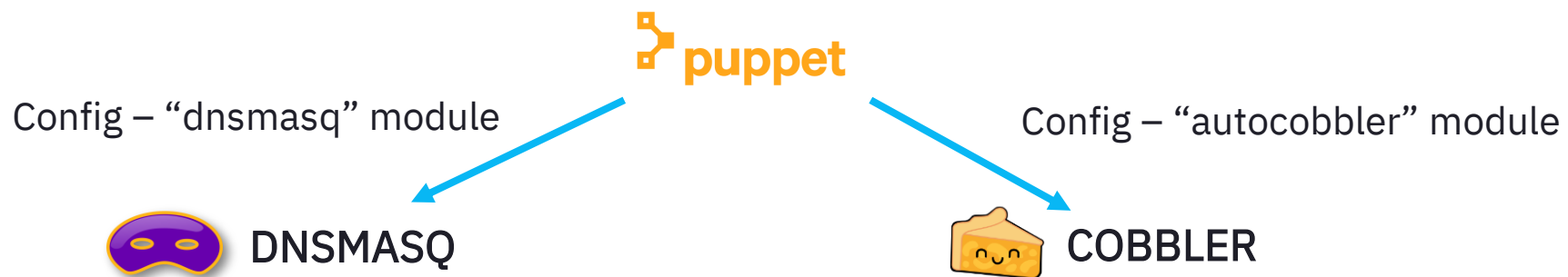
2024-04-11 18:45:45,274 PATCH method returned 200 status code, all is fine / Managers/System.Embedded.1/Action
Hot Spare parameter set to Disabled

=== [172.29.34.10] Step 5 of 7: Set BIOS parameters ===

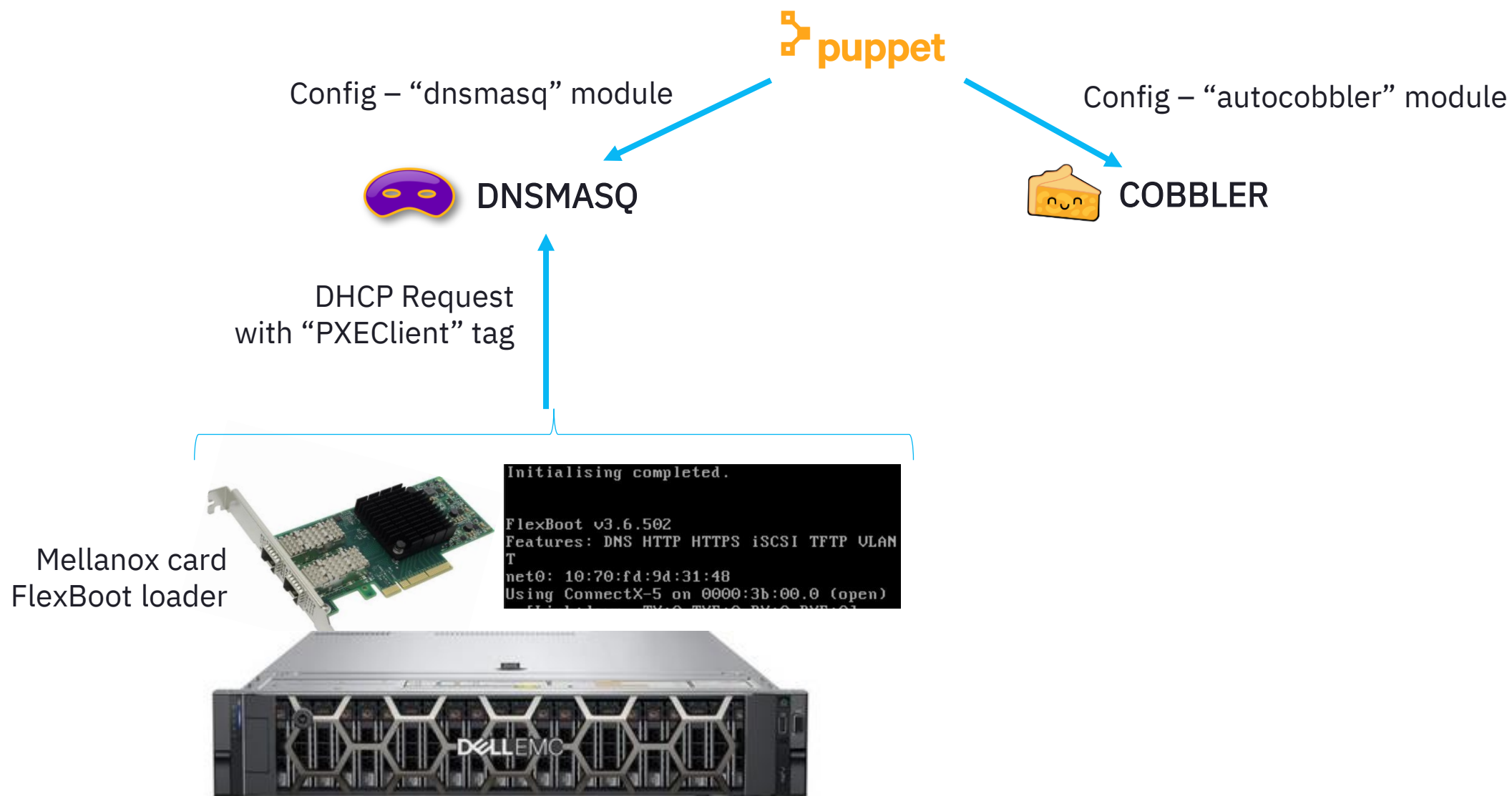
2024-04-11 18:45:45,584 POST method returned 200 status code, all is fine / Systems/System.Embedded.1/Bios/
2024-04-11 18:45:48,653 PATCH method returned 200 status code, all is fine / Systems/System.Embedded.1/Bios/
2024-04-11 18:45:53,412 POST method returned 200 status code, all is fine / Managers/iDRAC.Embedded.1/Job
2024-04-11 18:46:14,013 POST method returned 204 status code, all is fine / Systems/System.Embedded.1/Action
2024-04-11 18:50:34,313 Current job state is not done, will retry in 1 min (attempt 1/15)
2024-04-11 18:51:34,841 Current job state is not done, will retry in 1 min (attempt 2/15)
2024-04-11 18:52:35,205 Current job state is not done, will retry in 1 min (attempt 3/15)
```

4. Разливка ОС

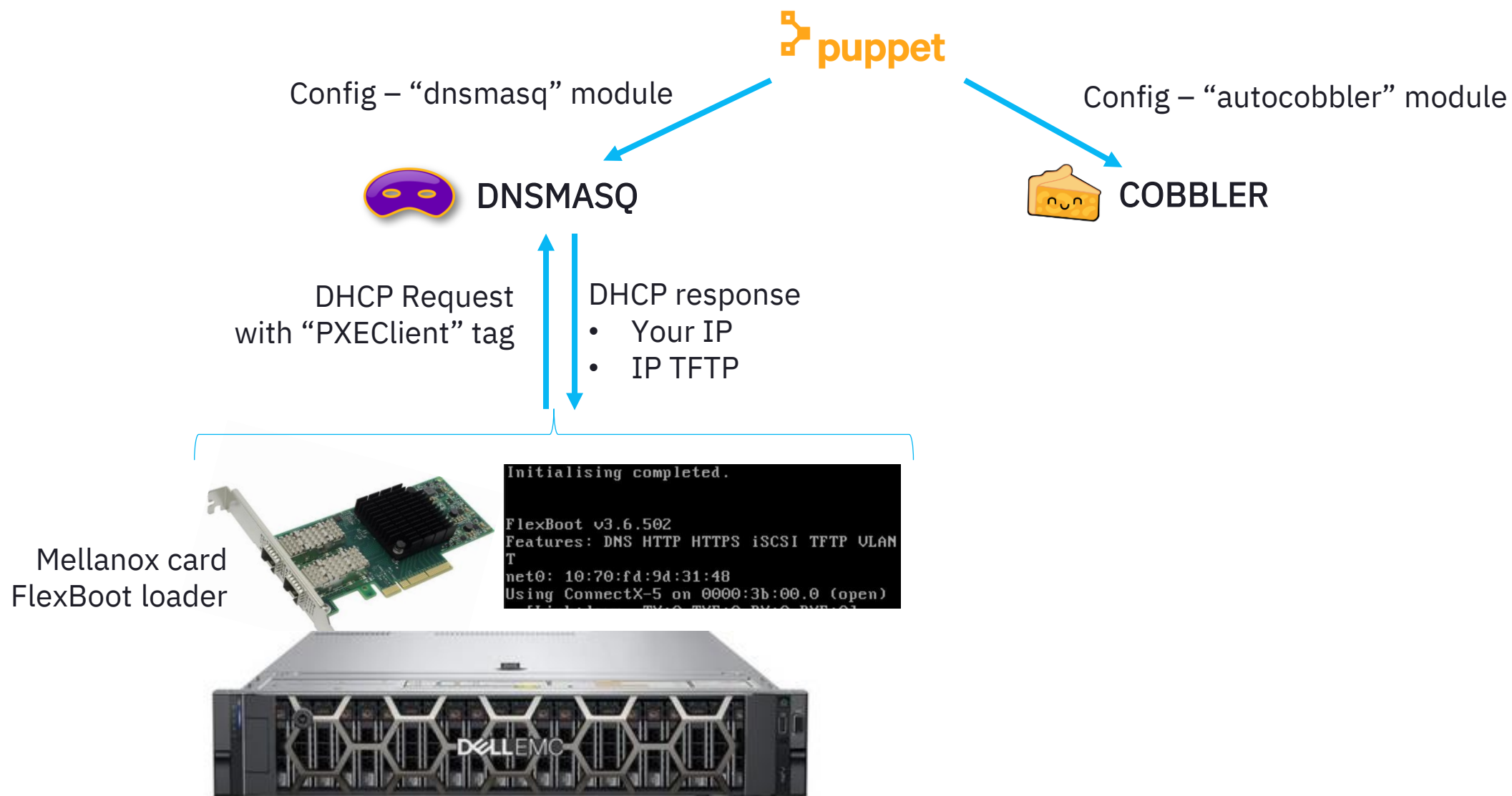
4. Разливка ОС



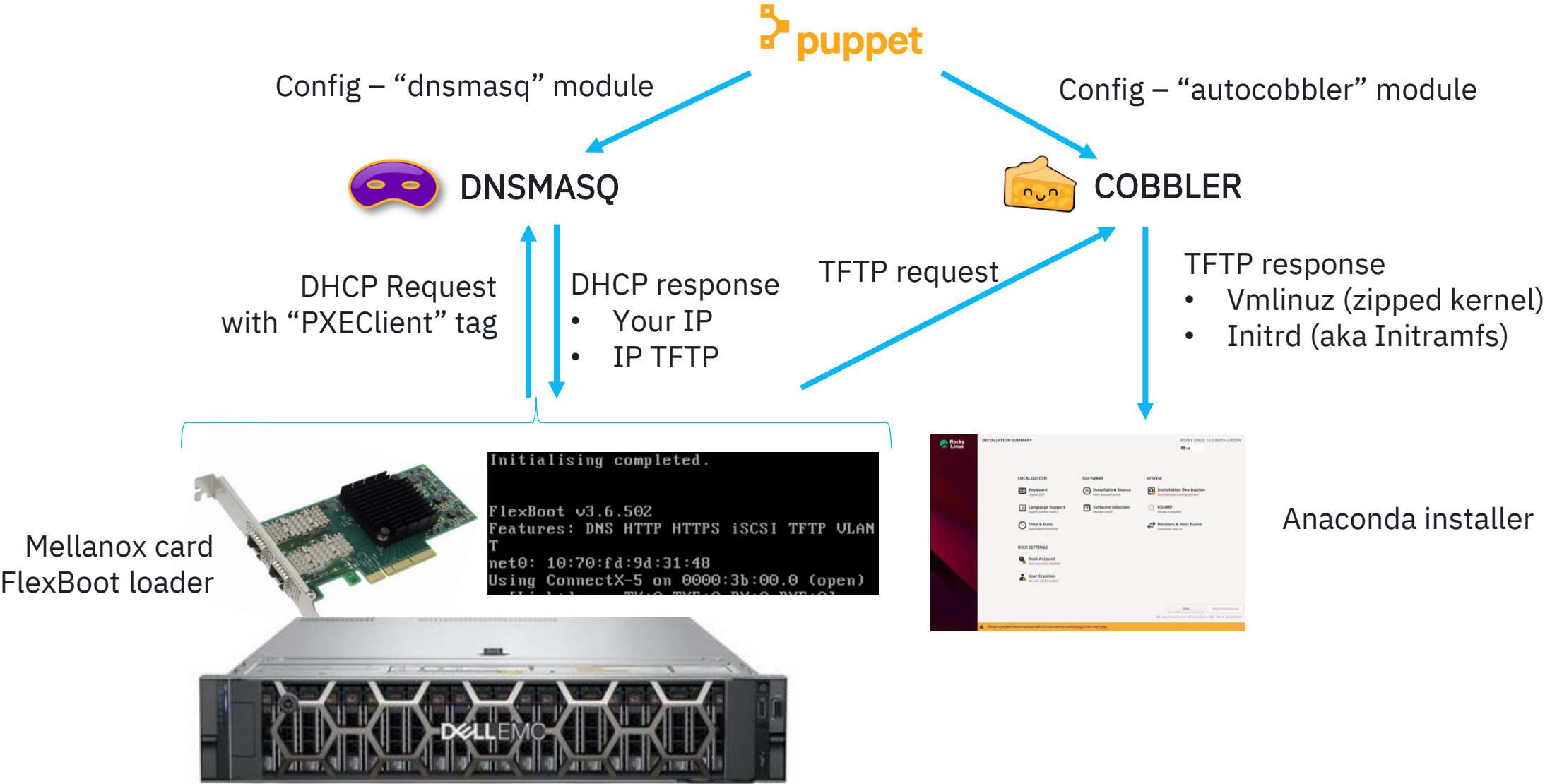
4. Разливка ОС



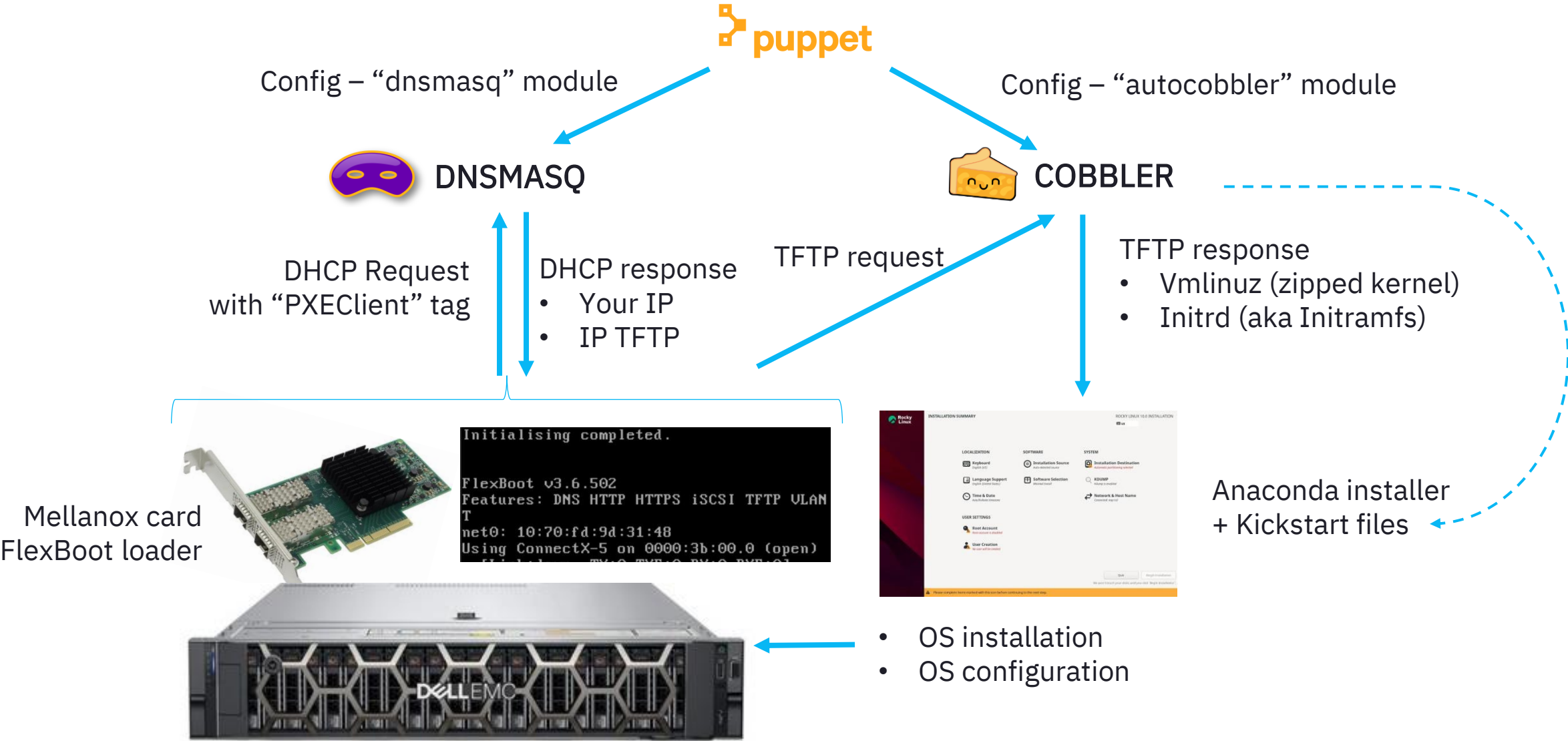
4. Разливка ОС



4. Разливка ОС



4. Разливка ОС



4. Разливка ОС

Фрагмент Kickstart шаблона из модуля autocoobler

```
1 #####
2 ## Kickstart RHEL8
3 #####
4 # Default language for system
5 lang en_US
6 # Default language for keyboard
7 keyboard us
8 # Default language for timezone
9 timezone Europe/Moscow --isUtc
10 [REDACTED]
11 #platform x86, AMD64, or Intel EM64T
12 url --url=$tree
13
14 # Use only /dev/sda device for not to clear any other devices with possible data
15 [REDACTED]
16 ignoredisk --only-use=sda
17
18 # If any cobbler repo definitions were referenced in the kickstart profile, include them here.
19 $yum_repo_stanza
20 reboot
21 bootloader --location=mbr --append="rhgb quiet crashkernel=0M-2G:128M,2G-6G:256M,6G-12G:512M,12G-64G:768M,64G-:1G biosdevname=0 net.ifnames=0 bonding.max_bonds=0"
22 zerombr
23 # Use only /dev/sda device for not to clear any other devices with possible data
24 [REDACTED]
25 clearpart --drives=sda --all --initlabel
26 volgroup c2cloud --pesize=4096 pv.0
27 # Create the first partition for BOOT
28 part /boot --fstype="xfs" --ondisk=sda --size=512
29 # required partition for UEFI
30 part /boot/efi --fstype vfat --ondisk=sda --size=128
31 # Create the secont partition for LVM with all other partitions
32 part pv.0 --fstype=lvm pv --ondisk=sda --size=21200 --grow
```

4. Разливка ОС



Сервер после разливки ОС

- Правильно размечен диск
- Установлены дополнительные пакеты (например puppet-agent)
- Базово сконфигурирована сеть
- Сконфигурирован доступ по SSH

Сервер сразу готов к дальнейшему деплою сервисов

5. Ввод в облако – утилита node bootstrap

Что делает:

- Выполняет предстартовые проверки сервера
- Запускает puppet
- Деплоит конфигури Облака
- Деплоит и запускает сервисы Облака на сервере

Умеет:

- Автоматически загружать параметры настройки сервера
- Перезагружать сервер и автоматически возобновлять работу с ним
- Запускать puppet на других узлах

```
[23:54]: ~]# c2-admin-node bootstrap run 20250722_144251_7148da8 sr101 --ccus 4.46 --g
oup gen1
Detected host roles: nc.
Selected host role: nc.
> ssh sr101 'cat /etc/redhat-release'
Detected OS version: CentOS 8.4.
> ssh sr101 'lscpu | grep \'^Model name:\''
Detected CPU model: Intel(R) Xeon(R) CPU E5-2440 0 @ 2.40GHz.
Detected puppet profile: 4.18.0-477.27.1.el8_8-v1.
2025.08.24 23:55:21.359 ( base.py:0272): [admin.scripting.script] D: Configuring Raven for host:
Block 0: NC node deploy (CCUs: 4.46, node group: gen1, dedicated: False)
Do you want to continue [y/n]: y
Step 0: Check full hostname and validate it in the deployment configuration
> ssh sr101 hostname -f
sr101.sr.123.ops
> Executing check_hostname(sr101
> Executing check_deployment(sr101)
Step 1: Clear old ssh keys from known hosts
> Executing get_host_address(sr101)
> c2-pdsh admin -- ssh-keygen -R sr101
```

6. Проверка - утилита c2-check

- Выполняет проверки, набор которых зависит от роли и версии ОС
- Печатает итоговый результат в виде таблицы “Ok / Not ok”
- Никак не влияет на работу проверяемого сервера и не требует root

```
18:39]:[ # c2-check nc sf025
Detected OS version: CentOS
100.0% | Checking OVS protocols
Status: FAIL
```

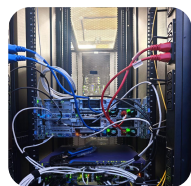
check	expected	current	status
Hostname	sf025	sf025	OK
OFED packages	ofed-scripts mlnx-ofa_kei mlnx-ofa_kei	ofed-scripts mlnx-ofa_kei mlnx-ofa_kei	OK
OVS packages	openvswitch	openvswitch	OK
OVS-KMOD packages	openvswitch	package oper	FAIL
OVN packages	ovn-22.09.1 ovn-host-22	package ovn package ovn EMC-ScaleIO	FAIL
SDC packages	EMC-ScaleIO	EMC-ScaleIO	OK
kdump packages	kexec-tools	kexec-tools	OK
BIOS version	1.8.2	1.8.2	OK
iDRAC version	6.10	6.10	OK
CPU stepping	6	6	OK
Memory frequency (MHz)	3200	3200	OK
kdump memory	141557760	1073741824	OK
kdump service	active	active	OK
puppet service	active	active	OK
puppet node profile			OK
Resolve DNS	True	True	OK
hyperthreading	enabled	enabled	OK
c2roles.yaml	nc	nc	OK
Kdump config	ssh root@kd	ssh root@kd	OK
Kernel version			OK
c2-kd scripts in initramfs	True	True	OK
hugepages	1522729	1522729	OK
Cluster speed	25000	25000	OK
Clusterbackup speed	1000	1000	OK
Cluster interfaces state	ok	ok	OK
Network access tests	All hosts are available	All hosts are available	OK
MTU value	{'cluster': 9170, 'storage': 9170}	{'cluster': 9170, 'storage': 9170}	OK
OVS bridge exists	br-int	br-int	OK
OVN registration	Correct	Correct	OK
qemu ignore_msr flag	Y	Y	OK
OVS protocols	Correct	Correct	OK

Этапы ввода нового сервера в Облако

1

Определение места
установки

- opendcim-cli
- c2-prepare-host units



4

Разливка ОС

- Cobbler + “Autocobbler”
puppet module



2

Подготовка и выкатка
конфигов

- Локальные скрипты
- Auto DNS/Ansible/Hiera



5

Ввод в облако

- c2-admin-node bootstrap



3

Подготовка сервера
c2-prepare-host



6

Проверка

- c2-check



Планируется объединение в единый сервис...

Выводы

Выводы

- Стандартизация инфраструктуры — не бюрократия, а инвестиция, которая окупается при каждом новом развёртывании

Выводы

- Стандартизация инфраструктуры — не бюрократия, а инвестиция, которая окупается при каждом новом развёртывании
- Спланировать на начальном этапе проще и дешевле, чем переделывать под нагрузкой:
 - Физическое размещение и энергопотребление железа
 - Топологию сетей и количество коммутации

Выводы

- Стандартизация инфраструктуры — не бюрократия, а инвестиция, которая окупается при каждом новом развёртывании
- Спланировать на начальном этапе проще и дешевле, чем переделывать под нагрузкой:
 - Физическое размещение и энергопотребление железа
 - Топологию сетей и количество коммутации
- Нельзя полностью автоматизировать работу с железными серверами.
Но можно сделать этот процесс быстрым и комфортным для инженеров, используя:
 - API DCIM систем
 - Redfish API
 - Разливку ОС по PXE в связке с системами управления конфигурациями



Спасибо за внимание

<https://k2.cloud>
info@k2.cloud

Как мы поддерживаем
«бесконечность» Облака

