# Dokumentacja powykonawcza

**Klastra obliczeniowego dla modelu Aladin w Instytucie Meteorologii i Gospodarki Wodnej**

# Spis treści

# 1. Specyfikacja sprzętu

## 1.1. Obudowa kasetowa dla serwerów obliczeniowych (chassis HPC) – 6 szt.

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | 507019-B21 | HP BLc7000 CTO 3 IN LCD ROHS Encl |
| 1 | 438030-B21 | HP BLc GbE2c LY 2/3 Switch |
| 1 | 489184-B21 | HP BLc 4X QDR IB Switch |
| 1 | 517521-B21 | HP 6X 2400W Gold Ht Plg FIO Pwr Sply Kit |
| 1 | 413381-B21 | HP BLc7000 3 PH Intl FIO Power Mod Opt |
| 1 | 517520-B21 | HP BLc 6X Active Cool 200 FIO Fan Opt |
| 1 | 436670-B21 | HP C-Class Blade IB Cable Bracket |

## 1.2. Serwer obliczeniowy (N1-N97) – 96 szt.

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | 641016-B21 | HP BL460c Gen8 10Gb FLB CTO Blade |
| 1 | 662076-L21 | HP BL460c Gen8 E5-2690 FIO Kit |
| 1 | 662076-B21 | HP BL460c Gen8 E5-2690 Kit |
| 8 | 672631-B21 | HP 16GB 2Rx4 PC3-12800R-11 Kit |
| 1 | 684212-B21 | HP FlexFabric 10Gb 2P 554FLB FIO Adptr |
| 1 | 644160-B21 | HP IB QDR/EN 10Gb 2P 544M Adptr |

## 1.3. Obudowa kasetowa dla serwerów zarządzających (chassis MGMT) – 1 szt.

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | 507019-B21 | HP BLc7000 CTO 3 IN LCD ROHS Encl |
| 2 | AJ821B | HP B-series 8/24c BladeSystem SAN Switch |
| 1 | 516733-B21 | HP 6120XG Blade Switch |
| 1 | 489184-B21 | HP BLc 4X QDR IB Switch |
| 4 | AJ716B | HP 8Gb Short Wave B-Series SFP+ 1 Pack |
| 2 | 453154-B21 | HP BLc VC 1Gb RJ-45 SFP Opt Kit |
| 2 | 455889-B21 | HP BLc 10Gb LRM SFP+ Opt |
| 1 | 517521-B21 | HP 6X 2400W Gold Ht Plg FIO Pwr Sply Kit |
| 1 | 456204-B21 | HP BLc7000 DDR2 Encl Mgmt Option |
| 1 | 413381-B21 | HP BLc7000 3 PH Intl FIO Power Mod Opt |
| 1 | 517520-B21 | HP BLc 6X Active Cool 200 FIO Fan Opt |
| 1 | 436670-B21 | HP C-Class Blade IB Cable Bracket |

## 1.4. Węzły zarządzające i LUSTRE - MASTER1, MASTER2, OSS1, OSS2, MDS1, MDS2

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | 641016-B21 | HP BL460c Gen8 10Gb FLB CTO Blade |
| 1 | 662064-L21 | HP BL460c Gen8 E5-2670 FIO Kit |
| 1 | 662064-B21 | HP BL460c Gen8 E5-2670 Kit |
| 4 | 672631-B21 | HP 16GB 2Rx4 PC3-12800R-11 Kit |
| 2 | 652564-B21 | HP 300GB 6G SAS 10K 2.5in SC ENT HDD |
| 1 | 684212-B21 | HP FlexFabric 10Gb 2P 554FLB FIO Adptr |
| 1 | 644160-B21 | HP IB QDR/EN 10Gb 2P 544M Adptr |

| | | |
|---|---|---|
| 1 | 651281-B21 | HP QMH2572 8Gb FC HBA |

## 1.5. Węzeł zarządzający MASTER3

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | 727021-B21 | HP BL460c Gen9 10Gb/20Gb FLB CTO Blade |
| 1 | 726991-L21 | HP BL460c Gen9 E5-2650v3 FIO Kit |
| 1 | 726991-B21 | HP BL460c Gen9 E5-2650v3 Kit |
| 4 | 726719-B21 | HP 16GB 2Rx4 PC4-2133P-R Kit |
| 2 | 652564-B21 | HP 300GB 6G SAS 10K 2.5in SC ENT HDD |
| 1 | 766491-B21 | HP FlexFabric 10Gb 2P 536FLB FIO Adptr |
| 1 | 710608-B21 | HP QMH2672 16Gb FC HBA |
| 1 | 764282-B21 | HP IB QDR/EN 10Gb 2P 544+M Adptr |
| 1 | 761878-B21 | HP H244br FIO Smart HBA |

## 1.6. Przełącznik InifniBand zewnętrzny (IB SW) z okablowaniem

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 4 | 712495-B21 | Mellanox IB QDR/FDR10 36P Switch |
| 16 | 498385-B21 | HP 1M 4X DDR/QDR QSFP IB Cu Cable |
| 16 | 498385-B22 | HP 2M 4X DDR/QDR QSFP IB Cu Cable |
| 64 | 498385-B23 | HP 3M 4X DDR/QDR QSFP IB Cu Cable |
| 12 | 498385-B24 | HP 5M 4X DDR/QDR QSFP IB Cu Cable |

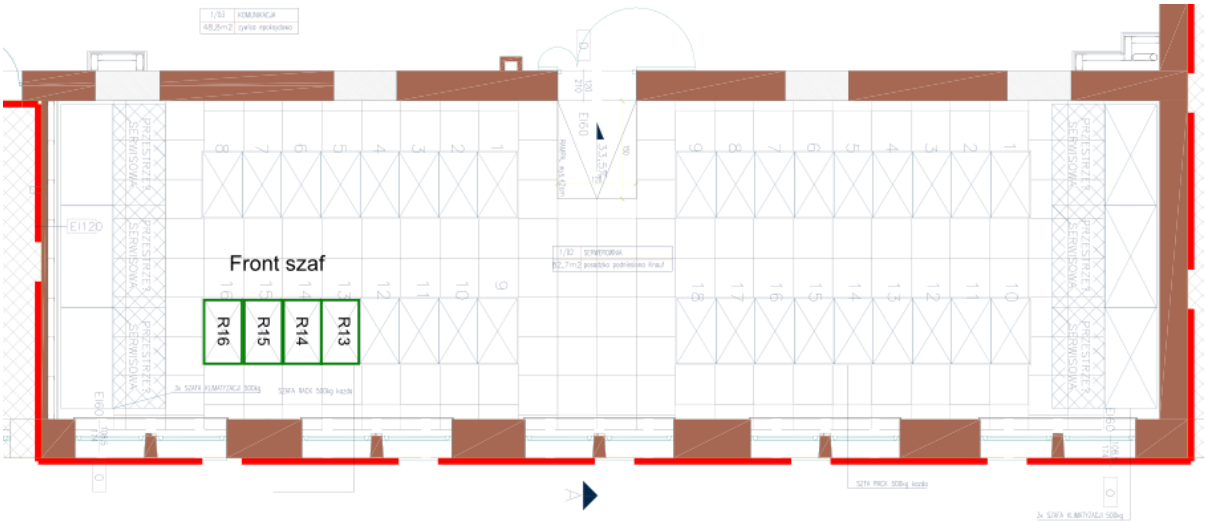
## 1.7. Macierz dyskowa

| Szt | P/N | Nazwa podzespołu |
|---|---|---|
| 1 | K2R80A | HPE MSA 2040 Energy Star SAN Dual Controller SFF Storage |
| 2 | C8R23A | HPE MSA 2040 8Gb Short Wave Fibre Channel SFP+ 4-pack Transceiver |
| 7 | N9X95A | HPE MSA 400GB 12G SAS MU 2.5in SSD |
| 4 | M0S96A | HPE MSA 2040 Energy Star LFF Disk Enclosure |
| 41 | AW555A | HP P2000 2TB 6G SAS 7.2K rpm LFF (3.5-inch) Dual Port MDL Hard Drive |

# 2. Instalacja sprzętu

## 2.1. Instalacja szaf teleinformatycznych

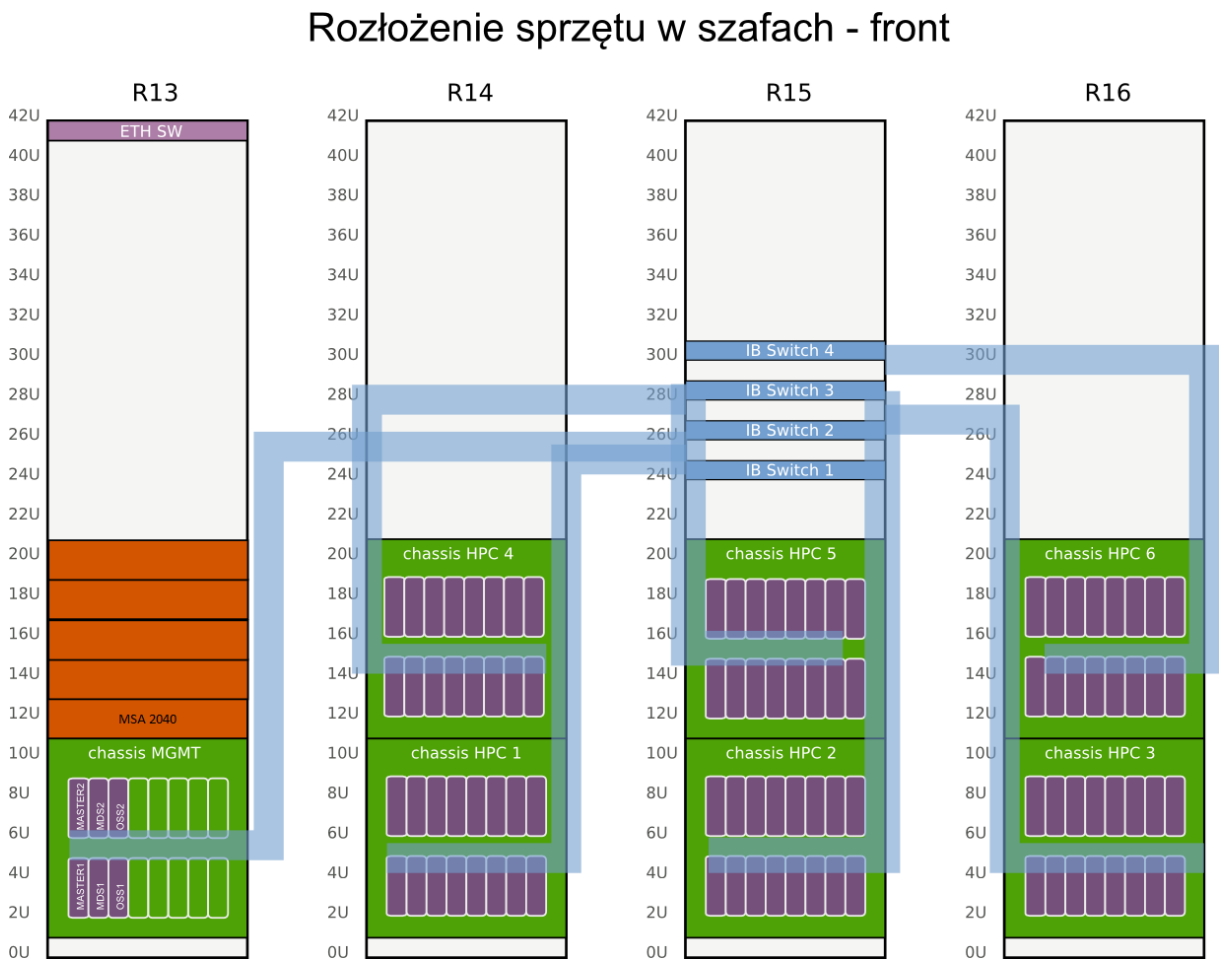
Klaster został zamontowany w 4 szafach BKT WEST B o numerze katalogowym 11038613.2V3DB o wysokości 42U. Szafy zostały umieszczone w serwerowni jak na rysunku




**Rysunek 1. Rozmieszczenie szaf w serwerowni**

## 2.2. Instalacja sprzętu w szafach

Całość sprzętu wchodząca w skład klastra obliczeniowego została zamontowana w szafach. Rozkład urządzeń został przedstawiony na rysunku poniżej.




**Rysunek 2. Rozmieszczenie sprzętu w szafie**

### 2.2.1. Zasilanie

Każda obudowa kasetowa C7000 podłączona została za pomocą dwóch wtyków IEC-309 200/346 V - 240/415 V, 5-Pin, 6h 16 A, które zostały podłączone do odpowiednich gniazd wyprowadzonych z kaset odpływowych. Szafy wyposażone są w kasety odpływowe z gniazdami:

1. R13
    - 2 gniazda IEC-309 200/346 V - 240/415 V, 5-Pin, 6h 16 A
    - 2 gniazda schuko 16A do podłączenia PDU
2. R14
    - 4 gniazda IEC-309 200/346 V - 240/415 V, 5-Pin, 6h 16 A
3. R15
    - 4 gniazda IEC-309 200/346 V - 240/415 V, 5-Pin, 6h 16 A
    - 2 gniazda schuko 16A do podłączenia PDU
4. R16
    - 4 gniazda IEC-309 200/346 V - 240/415 V, 5-Pin, 6h 16 A

W szafie R13 i R15 zainstalowane zostały listwy zasilające (PDU) do zasilenia przełączników oraz macierzy dyskowej.

## 2.3. Sprzęt serwerowy

Serwery obliczeniowe zainstalowane zostały w obudowach kasetowych HP c7000. Znajdujące się w nich serwery połówkowe blade BL460c Gen8 realizują funkcjonalność węzłów obliczeniowych klastra obliczeniowego.

## 2.4. Oznaczenie sprzętu

Węzły klastra posiadają nazwy w formacie n<n>, gdzie <n> jest numerem kolejnym węzła. W celu ułatwienia zarządzania klastrem nadano odpowiednie nazwy węzłom w systemach operacyjnych (hostname) przypisane do adresów IP. Dodatkowo na każdym urządzeniu przyklejona została naklejka z opisem.

# 3. Konfiguracja sieciowa

## 3.1. Sieci w klastrze

W klastrze dostępne są następujące sieci:

- Sieć obliczeniowa (IB)
- Sieć zarządzająca (MGMT)
- Sieć wewnętrzna klastra (HPC)
- Sieć dostępowa (sieć IMGW)

W tabeli poniżej przedstawiono adresację dla sieci:

| Nazwa | Adres sieci | Brama domyślna | Opis |
|---|---|---|---|
| HPC | 192.168.4.0/22 | 192.168.4.1 | Sieć wewnętrzna klastra |
| IB | 192.168.12.0/22 | 192.168.12.1 | Sieć obliczeniowa |
| MGMT | 172.31.71.0/24 | 172.31.71.1 | Sieć zarządzająca |

**Tabela 1. Opis sieci systemowych**

Adresy zostały dobrane tak, aby nie kolidowały z adresacją wewnętrzną IMGW.

## 3.2. Podłączenia Ethernet i VLAN

Każdej z sieci został przypisany odpowiadający VLAN skonfigurowany na przełącznikach.

Na przełącznikach sieciowych zostały skonfigurowane VLANy:

- Na przełączniku zewnętrznym:
  - [1-8] VLAN MGMT – podłączenie interfejsów OA obudów C7000
  - [9-10] VLAN MGMT – podłączenie kontrolerów macierzy MSA 2040
  - [11-17] VLAN MGMT – podłączenie przełączników 1GbE w obudowach C7000
  - [18] VLAN MGMT – podłączenie przełączników FC w obudowie C7000
- Na przełącznikach 1GbE w obudowach chassis HPC[1-6]
  - [1-16] VLAN HPC
  - [20-24] VLAN HPC
- Na przełączniku 1GbE w obudowie chassis MGMT
  - [1-3,9-11,17,22] VLAN HPC oraz tagowane VLAN MGMT
  - [18-19] Porty zewn. 10GbE – VLAN IMGW

Poprawne działanie klastra wymaga dostępu z VLAN HPC do zewnętrznej usługi katalogowej IMGW.

### 3.3. Sieć zarządzająca - MGMT

Do sieci zarządzającej przyłączone są interfejsy administracyjne takie jak interfejsy modułów zarządzających serwerów, przełączników sieciowych, macierzy. Interfejsy te zostały oddzielone od sieci użytkowników klastra tak, aby użytkownicy nie mieli dostępu do zarządzania sprzętem. Poniżej przedstawiono adresację urządzeń w sieci zarządzającej:

| Nazwa | Adres IP |
|---|---|
| OA MGMT A | 172.31.71.110 |
| OA HPC1 | 172.31.71.111 |
| OA HPC2 | 172.31.71.112 |
| OA HPC3 | 172.31.71.113 |
| OA HPC4 | 172.31.71.114 |
| OA HPC5 | 172.31.71.115 |
| OA HPC6 | 172.31.71.116 |
| OA MGMT B | 172.31.71.117 |
| MSA 2040 | 172.31.71.129<br>172.31.71.130 |
| SW MGMT | 172.31.71.120 |
| SW HPC1 | 172.31.71.121 |
| SW HPC2 | 172.31.71.122 |
| SW HPC3 | 172.31.71.123 |
| SW HPC4 | 172.31.71.124 |
| SW HPC5 | 172.31.71.125 |
| SW HPC6 | 172.31.71.126 |
| Switch FC1 | 172.31.71.127 |
| Switch FC2 | 172.31.71.128 |
| master1 | 172.31.71.131 |
| master2 | 172.31.71.132 |
| mds1 A | 172.31.71.133 |
| mds2 A | 172.31.71.134 |

| | |
|---|---|
| oss1 A | 172.31.71.135 |
| oss2 A | 172.31.71.136 |
| n1 – n96 | 172.31.71.141-236 |
| manager OS | 172.31.71.248 |
| master1 OS | 172.31.71.249 |
| master2 OS | 172.31.71.250 |
| mds1 OS | 172.31.71.251 |
| mds2 OS | 172.31.71.252 |
| oss1 OS | 172.31.71.253 |
| oss2 OS | 172.31.71.254 |

**Tabela 2. Adresacja zarządzania fizycznego**

### 3.4. Adresacja IP sieci

Poniżej przedstawiono adresy IP skonfigurowane w systemach operacyjnych węzłów klastra:

| | Sieć | | |
|---|---|---|---|
| **Host** | HPC | IB | MGMT |
| n1 – n96 | 192.168.5.[1-96] | 192.168.15.[1-96] | - |
| master1 | 192.168.4.11 | 192.168.14.11 | 172.31.71.249 |
| master2 | 192.168.4.12 | 192.168.14.12 | 172.31.71.250 |
| mds1 | 192.168.4.21 | 192.168.14.21 | 172.31.71.251 |
| mds2 | 192.168.4.22 | 192.168.14.22 | 172.31.71.252 |
| oss1 | 192.168.4.31 | 192.168.14.31 | 172.31.71.253 |
| oss2 | 192.168.4.32 | 192.168.14.32 | 172.31.71.254 |
| headnode | 192.168.4.50 | 192.168.14.50 | 172.31.71.246 |
| nethost | 192.168.4.51 | 192.168.14.51 | - |
| license | 192.168.4.53 | 192.168.14.53 | - |

**Tabela 3. Wykaz adresacji IP systemów operacyjnych**

# 4. Konfiguracja pamięci masowej

## 4.1. Połączenia w sieci SAN

Do komunikacji serwerów zarządzających MASTER oraz serwerów LUSTRE z macierzą MSA 2040 zostały wykorzystane przełączniki FC w obudowie kasetowej C7000. Podłączenie przełączników do macierzy zostało wykonane za pomocą redundantnych połączeń FC.

Na przełącznikach został skonfigurowany odpowiedni zoning, tak aby zapewnić dostęp do zasobów dyskowych serwerom MASTER, MDS, OSS.

## 4.2. Konfiguracja macierzy MSA 2040

Macierz wyposażona jest w:

- 41 szt. dysków 2TB SAS 7.2K, z których zostały utworzone 4 grupy RAID6 8D+2P, jeden dysk został użyty jako hot-spare
- 7 szt. dysków 400GB SAS SSD, z których zostały utworzone 2 grupy RAID5 2D+1P, jeden dysk został użyty jako hot-spare

Na każdej grupie RAID z dysków 2TB został utworzony wolumen logiczny o pojemności całej grupy, 4 powstałe wolumeny zostały użyte jako zasoby OST dla LUSTRE i oznaczone w systemach jako: OST101, OST102, OST201, OST202. Zasoby OST zostały zgrupowane pod LUSTRE jako dwie niezależne pule dyskowe.

Na grupach dyskowych złożonych z dysków SSD zostały utworzone bezpośrednio wolumeny dyskowe (LUN):

1. MDT o rozmiarze 300GiB wykorzystany jako LUSTRE MDT
2. MGT o rozmiarze 1GiB wykorzystany jako LUSTRE MGT
3. Reszta przestrzeni dyskowej została przydzielona na LUN-y dla maszyn wirtualnych

W systemie operacyjnym węzłów MASTER, OSS i MDS został użyty mechanizm multipath do podłączenia zasobów dyskowych i zapewnienia wysokiej niezawodności. Za pomocą mechanizmu multipath zostały przydzielone również odpowiednie nazwy urządzeniom dyskowym w katalogu /dev/, aby możliwa była ich łatwa identyfikacja.

# 5. Serwery zarządzające i serwery wirtualne

Na serwerach MASTER1 i MASTER2 zainstalowany został system operacyjny zgodny z RHEL6 posiadający wsparcie dla wirtualizacji KVM.

System operacyjny został zainstalowany na dyskach lokalnych po wcześniejszym utworzeniu sprzętowego zabezpieczenia RAID na tych dyskach.

Na serwerach MASTER interfejsy sieciowe zostały skonfigurowane z użyciem mechanizmu VLAN tak, że uzyskane zostaną 3 interfejsy z każdej z sieci:

- HPC
- MGMT
- IMGW

Interfejsy zostały połączone z odpowiadającymi im mostami sieciowymi (bridge), które będą mogły być użyte do podłączenia maszyn wirtualnych do odpowiednich sieci klastra.

## 5.1. Maszyny wirtualne na serwerach MASTER

Zostały skonfigurowane następujące maszyny wirtualne:

| Host | Opis |
|---|---|
| headnode | Węzeł dostępowy oraz menadżer zadań/kolejkowania |
| nethost | Serwer usług sieciowych (wewnętrzne usługi DNS, DHCP, TFTP) |
| license | Serwer sieciowych licencji oprogramowania |

Na każdej z maszyn wirtualnych został zainstalowany system zgodny z RHEL6 oraz odpowiednie oprogramowanie.

## 5.2. Klaster wysokiej dostępności

Na maszynach MASTER, MDS, OSS został skonfigurowany klaster wysokiej dostępności (HA), aby zapewnić wysoką dostępność.

Skonfigurowany został mechanizm Fencing, na interfejsach zarządzających (OA) obudów kasetowych chassis MGMT utworzony został użytkownik z wygenerowanym losowo hasłem służący do restartowania maszyn fizycznych przez klaster HA.

## 5.3. System składowania danych - LUSTRE

System plików LUSTRE został zainstalowany na serwerach MDS oraz OSS. Użyta została stabilna wersja LUSTRE – 2.1.6

System LUSTRE składa się z zasobów metadanych (MDT) i danych (OST). Zasoby te zostały udostępnione z macierzy na serwery MDS i OSS. Zasoby są udostępnione z serwerów zgodnie ze schematem:

- Serwer MDS – LUNy: MGT, MDT
- Serwer OSS1 - LUNy: OST101, OST102
- Serwer OSS2 - LUNy: OST201, OST202
- Na serwerach MDS został uruchomiony zasób MGT oraz MDT z macierzy. System plików został wykreowany z nazwą klastra obliczeniowego.

Na serwerach OSS1 i OSS2 został skonfigurowany klaster HA, którego zadaniem jest przełączanie zasobów LUSTRE na działający serwer OSS. Klienci LUSTRE zostali skonfigurowani tak, że w przypadku awarii i przełączenia zasobu OST na inny serwer OSS lub przełączenia zasobu MDT na inny serwer MDS operacje plikowe będą kontynuowane.

Całość komunikacji LUSTRE odbywa się z natywnym użyciem sieci InfiniBand za pomocą LUSTRE o2ib.

Rozbudowa systemu składowania danych może się odbywać poprzez dodawanie nowych dysków, macierzy oraz serwerów OSS.

## 5.4. Menadżer kolejkowania (zasobów)

Został skonfigurowany menadżer kolejkowania Torque w wersji 2.5 w wersji przygotowanej do współpracy z narzędziami do zarządzania.

W menadżerze zasobów zostały skonfigurowane kolejki testowa i produkcyjna. Zostały też przygotowane odpowiednie uprawnienia użytkowników do dodawania zadań do kolejek.

## 5.5. Serwery obliczeniowe

Serwery obliczeniowe zostały przygotowane do uruchamiania za pomocą mechanizmu PXE oraz systemu plików LUSTRE. Za pomocą dedykowanego mechanizmu jest uruchamiany system operacyjny bezpośrednio z systemu plików LUSTRE.

Modyfikacje obrazów systemu węzłów oraz instalacja oprogramowania odbywa się za pomocą mechanizmu chroot z serwerów zarządzających lub serwera dostępowgo – headnode.

## 5.6. Skrypty uruchomieniowe modelu ALADIN

W ramach prac został wykonany projekt wdrożenia skryptów obliczeniowych zgodnie z założeniami:

- Skrypty uruchomieniowe muszą posiadać flagę rerunnable i mieć możliwość automatycznego restartu po awarii oznacza to, że każdy skrypt musi być parametryzowany co do parametrów

brzegowych oraz położenia plików, a także tworzyć przed uruchomieniem środowisko plikowe (kopiowanie plików przed obliczeniami i po obliczeniach).

- Jako katalog roboczy obliczeń należy wykorzystać katalog zdefiniowany przez zmienną TMPDIR menadżera kolejkowania.
- Skrypty sprawdzające obecność danych powinny zostać zrealizowane jako zadania menadżera kolejkowania w trybie usługi (service), skrypt taki powinien być zgodny ze specyfikacją skryptów usług w systemach Linux LSB (Linux Standard Base).
- Zadania powinny sprawdzać dane wejściowe przed uruchomieniem oraz poprawność danych wyjściowych po uruchomieniu za pomocą standardowych narzędzi np. bibliotek do obsługi plików grib etc.
- Zadania powinny zostać przetestowane na możliwe awarie (np. brak miejsca na dysku, niekompletność plików wyjściowych) i zwracać poprawne kody wyjścia w przypadku prawidłowego zakończenia obliczeń oraz błędów.
- Zadania powinny ustawiać poprawne maksymalne wartości walltime (max. Czas działania), aby możliwe było wychwycenie sytuacji np. zawieszenia zadania obliczeniowego.

# 6. Uruchomione usługi

## 1.1 Serwery zarządzające – master

```
acpid              0:off 1:off 2:on  3:on  4:on  5:on  6:off
atd                0:off 1:off 2:off 3:on  4:on  5:on  6:off
auditd             0:off 1:off 2:on  3:on  4:on  5:on  6:off
autofs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
blk-availability   0:off 1:on  2:on  3:on  4:on  5:on  6:off
certmonger         0:off 1:off 2:off 3:on  4:on  5:on  6:off
cgconfig           0:off 1:off 2:on  3:on  4:on  5:on  6:off
cman               0:off 1:off 2:on  3:on  4:on  5:on  6:off
cpuspeed           0:off 1:on  2:off 3:off 4:off 5:off 6:off
crond              0:off 1:off 2:on  3:on  4:on  5:on  6:off
cups               0:off 1:off 2:on  3:on  4:on  5:on  6:off
haldaemon          0:off 1:off 2:off 3:on  4:on  5:on  6:off
ibacm              0:off 1:off 2:on  3:on  4:on  5:on  6:off
irqbalance         0:off 1:off 2:off 3:on  4:on  5:on  6:off
iscsi              0:off 1:off 2:off 3:on  4:on  5:on  6:off
iscsid             0:off 1:off 2:off 3:on  4:on  5:on  6:off
ksm                0:off 1:off 2:off 3:on  4:on  5:on  6:off
ksmtuned           0:off 1:off 2:off 3:on  4:on  5:on  6:off
libvirtd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
lvm2-monitor       0:off 1:on  2:on  3:on  4:on  5:on  6:off
mcelogd            0:off 1:off 2:off 3:on  4:off 5:on  6:off
mdmonitor          0:off 1:off 2:on  3:on  4:on  5:on  6:off
messagebus         0:off 1:off 2:on  3:on  4:on  5:on  6:off
multipathd         0:off 1:off 2:on  3:on  4:on  5:on  6:off
netfs              0:off 1:off 2:off 3:on  4:on  5:on  6:off
network            0:off 1:off 2:on  3:on  4:on  5:on  6:off
nfslock            0:off 1:off 2:off 3:on  4:on  5:on  6:off
ntpd               0:off 1:off 2:on  3:on  4:on  5:on  6:off
ntpdate            0:off 1:off 2:on  3:on  4:on  5:on  6:off
oddjobd            0:off 1:off 2:off 3:on  4:on  5:on  6:off
opensm             0:off 1:off 2:on  3:on  4:on  5:on  6:off
portreserve        0:off 1:off 2:on  3:on  4:on  5:on  6:off
postfix            0:off 1:off 2:on  3:on  4:on  5:on  6:off
rdma               0:off 1:off 2:on  3:on  4:on  5:on  6:off
rgmanager          0:off 1:off 2:on  3:on  4:on  5:on  6:off
ricci              0:off 1:off 2:on  3:on  4:on  5:on  6:off
rpcbind            0:off 1:off 2:on  3:on  4:on  5:on  6:off
rpcgssd            0:off 1:off 2:off 3:on  4:on  5:on  6:off
rpcidmapd          0:off 1:off 2:off 3:on  4:on  5:on  6:off
rsyslog            0:off 1:off 2:on  3:on  4:on  5:on  6:off
spice-vdagentd     0:off 1:off 2:off 3:off 4:off 5:on  6:off
sshd               0:off 1:off 2:on  3:on  4:on  5:on  6:off
sysstat            0:off 1:on  2:on  3:on  4:on  5:on  6:off
udev-post          0:off 1:on  2:on  3:on  4:on  5:on  6:off
```

### 1.2 Serwery zarządzające – mds

```
1  acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
2  atd               0:off 1:off 2:off 3:on  4:on  5:on  6:off
3  auditd            0:off 1:off 2:on  3:on  4:on  5:on  6:off
4  autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
5  avahi-daemon      0:off 1:off 2:off 3:on  4:on  5:on  6:off
6  blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
7  certmonger        0:off 1:off 2:off 3:on  4:on  5:on  6:off
8  cman              0:off 1:off 2:on  3:on  4:on  5:on  6:off
9  cpuspeed          0:off 1:on  2:off 3:off 4:off 5:off 6:off
10 crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
11 cups              0:off 1:off 2:on  3:on  4:on  5:on  6:off
12 haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
13 ibacm             0:off 1:off 2:on  3:on  4:on  5:on  6:off
14 irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
15 libvirt-guests    0:off 1:off 2:on  3:on  4:on  5:on  6:off
16 lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
17 mcelogd           0:off 1:off 2:off 3:on  4:off 5:on  6:off
18 mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
19 messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
20 multipathd        0:off 1:off 2:on  3:on  4:on  5:on  6:off
21 netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
22 network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
23 nfslock           0:off 1:off 2:off 3:on  4:on  5:on  6:off
24 ntpd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
25 ntpdate           0:off 1:off 2:on  3:on  4:on  5:on  6:off
26 oddjobd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
27 portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
28 postfix           0:off 1:off 2:on  3:on  4:on  5:on  6:off
29 rdma              0:off 1:off 2:on  3:on  4:on  5:on  6:off
30 rgmanager         0:off 1:off 2:on  3:on  4:on  5:on  6:off
31 ricci             0:off 1:off 2:on  3:on  4:on  5:on  6:off
32 rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
33 rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
34 rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
35 rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
36 spice-vdagentd    0:off 1:off 2:off 3:off 4:off 5:on  6:off
37 sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
38 sssd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
39 sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
40 udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
```

### 1.3 Serwery zarządzające – oss

```
1  acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
2  atd               0:off 1:off 2:off 3:on  4:on  5:on  6:off
3  auditd            0:off 1:off 2:on  3:on  4:on  5:on  6:off
4  autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
5  avahi-daemon      0:off 1:off 2:off 3:on  4:on  5:on  6:off
6  blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
7  certmonger        0:off 1:off 2:off 3:on  4:on  5:on  6:off
8  cman              0:off 1:off 2:on  3:on  4:on  5:on  6:off
9  cpuspeed          0:off 1:on  2:off 3:off 4:off 5:off 6:off
10 crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
11 cups              0:off 1:off 2:on  3:on  4:on  5:on  6:off
12 haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
13 ibacm             0:off 1:off 2:on  3:on  4:on  5:on  6:off
14 irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
15 libvirt-guests    0:off 1:off 2:on  3:on  4:on  5:on  6:off
16 lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
17 mcelogd           0:off 1:off 2:off 3:on  4:off 5:on  6:off
18 mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
19 messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
20 multipathd        0:off 1:off 2:on  3:on  4:on  5:on  6:off
21 netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
22 network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
23 nfslock           0:off 1:off 2:off 3:on  4:on  5:on  6:off
24 ntpd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
25 ntpdate           0:off 1:off 2:on  3:on  4:on  5:on  6:off
26 portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
27 postfix           0:off 1:off 2:on  3:on  4:on  5:on  6:off
28 rdma              0:off 1:off 2:on  3:on  4:on  5:on  6:off
29 rgmanager         0:off 1:off 2:on  3:on  4:on  5:on  6:off
30 ricci             0:off 1:off 2:on  3:on  4:on  5:on  6:off
31 rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
32 rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
33 rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
34 rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
35 spice-vdagentd    0:off 1:off 2:off 3:off 4:off 5:on  6:off
36 sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
37 sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
38 udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
```

### 1.4 Serwer usług sieciowych – nethost

```
1  acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
2  atd               0:off 1:off 2:off 3:off 4:on  5:on  6:off
3  auditd            0:off 1:off 2:on  3:off 4:on  5:on  6:off
4  autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
5  blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
6  certmonger        0:off 1:off 2:off 3:on  4:on  5:on  6:off
7  cpuspeed          0:off 1:on  2:on  3:off 4:on  5:on  6:off
8  crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
9  cups              0:off 1:off 2:on  3:off 4:on  5:on  6:off
10 dnsmasq           0:off 1:off 2:on  3:on  4:on  5:on  6:off
11 haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
12 irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
13 iscsi             0:off 1:off 2:off 3:off 4:on  5:on  6:off
14 iscsid            0:off 1:off 2:off 3:off 4:on  5:on  6:off
15 lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
16 mcelogd           0:off 1:off 2:off 3:off 4:off 5:on  6:off
17 mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
18 messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
19 netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
20 network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
21 nfs               0:off 1:off 2:on  3:off 4:on  5:on  6:off
22 nfslock           0:off 1:off 2:off 3:off 4:on  5:on  6:off
23 portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
24 postfix           0:off 1:off 2:on  3:on  4:on  5:on  6:off
25 rdma              0:off 1:off 2:on  3:off 4:on  5:on  6:off
26 rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
27 rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
28 rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
29 rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
30 spice-vdagentd    0:off 1:off 2:off 3:off 4:off 5:on  6:off
31 sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
32 sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
33 udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
34 vsftpd            0:off 1:off 2:on  3:on  4:on  5:on  6:off
```

### 1.5 Serwer dostępowy – headnode

```
1  acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
2  atd               0:off 1:off 2:off 3:on  4:on  5:on  6:off
3  auditd            0:off 1:off 2:on  3:on  4:on  5:on  6:off
4  autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
5  blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
6  certmonger        0:off 1:off 2:off 3:on  4:on  5:on  6:off
7  cpuspeed          0:off 1:on  2:on  3:on  4:on  5:on  6:off
8  crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
9  cups              0:off 1:off 2:on  3:on  4:on  5:on  6:off
10 haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
11 irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
12 kdump             0:off 1:off 2:off 3:on  4:on  5:on  6:off
13 lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
14 mcelogd           0:off 1:off 2:off 3:on  4:off 5:on  6:off
15 mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
16 messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
17 netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
18 network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
19 nfslock           0:off 1:off 2:off 3:on  4:on  5:on  6:off
20 ntpd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
21 oddjobd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
22 pbs_sched         0:off 1:off 2:on  3:on  4:on  5:on  6:off
23 pbs_server        0:off 1:off 2:on  3:on  4:on  5:on  6:off
24 portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
25 postfix           0:off 1:off 2:on  3:on  4:on  5:on  6:off
26 rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
27 rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
28 rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
29 rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
30 spice-vdagentd    0:off 1:off 2:off 3:off 4:off 5:on  6:off
31 sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
32 sssd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
33 sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
34 udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
```

### 1.6 Serwer licencji – license

```
1  acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
2  atd               0:off 1:off 2:off 3:off 4:on  5:on  6:off
3  auditd            0:off 1:off 2:on  3:off 4:on  5:on  6:off
4  autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
5  blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
6  certmonger        0:off 1:off 2:off 3:on  4:on  5:on  6:off
7  cpuspeed          0:off 1:on  2:on  3:on  4:on  5:on  6:off
8  crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
9  cups              0:off 1:off 2:on  3:off 4:on  5:on  6:off
10 haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
11 irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
12 kdump             0:off 1:off 2:off 3:off 4:on  5:on  6:off
13 lmgrd_intel       0:off 1:off 2:on  3:on  4:on  5:on  6:off
14 lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
15 mcelogd           0:off 1:off 2:off 3:off 4:off 5:on  6:off
16 mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
17 messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
18 netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
19 network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
20 nfslock           0:off 1:off 2:off 3:on  4:on  5:on  6:off
21 portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
22 postfix           0:off 1:off 2:on  3:off 4:on  5:on  6:off
23 rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
24 rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
25 rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
26 rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
27 sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
28 sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
29 udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
```

### 1.7 Węzły obliczeniowe

```
acpid             0:off 1:off 2:on  3:on  4:on  5:on  6:off
atd               0:off 1:off 2:off 3:on  4:on  5:on  6:off
autofs            0:off 1:off 2:off 3:on  4:on  5:on  6:off
blk-availability  0:off 1:on  2:on  3:on  4:on  5:on  6:off
cpuspeed          0:off 1:on  2:off 3:off 4:off 5:off 6:off
crond             0:off 1:off 2:on  3:on  4:on  5:on  6:off
haldaemon         0:off 1:off 2:off 3:on  4:on  5:on  6:off
ibacm             0:off 1:off 2:on  3:on  4:on  5:on  6:off
irqbalance        0:off 1:off 2:off 3:on  4:on  5:on  6:off
lvm2-monitor      0:off 1:on  2:on  3:on  4:on  5:on  6:off
mcelogd           0:off 1:off 2:off 3:on  4:off 5:on  6:off
mdmonitor         0:off 1:off 2:on  3:on  4:on  5:on  6:off
messagebus        0:off 1:off 2:on  3:on  4:on  5:on  6:off
netfs             0:off 1:off 2:off 3:on  4:on  5:on  6:off
network           0:off 1:off 2:on  3:on  4:on  5:on  6:off
nfslock           0:off 1:off 2:off 3:on  4:on  5:on  6:off
ntpd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
ntpdate           0:off 1:off 2:on  3:on  4:on  5:on  6:off
oddjobd           0:off 1:off 2:on  3:on  4:on  5:on  6:off
pbs_mom           0:off 1:off 2:on  3:on  4:on  5:on  6:off
portreserve       0:off 1:off 2:on  3:on  4:on  5:on  6:off
rpcbind           0:off 1:off 2:on  3:on  4:on  5:on  6:off
rpcgssd           0:off 1:off 2:off 3:on  4:on  5:on  6:off
rpcidmapd         0:off 1:off 2:off 3:on  4:on  5:on  6:off
rsyslog           0:off 1:off 2:on  3:on  4:on  5:on  6:off
sshd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
sssd              0:off 1:off 2:on  3:on  4:on  5:on  6:off
sysstat           0:off 1:on  2:on  3:on  4:on  5:on  6:off
udev-post         0:off 1:on  2:on  3:on  4:on  5:on  6:off
winbind           0:off 1:off 2:off 3:on  4:on  5:on  6:off
xinetd            0:off 1:off 2:off 3:on  4:on  5:on  6:off
```

# 7. Konfiguracja usług

## 1.8 Serwer zarządzający – master1

### 1.8.1 Usługa networking

- /etc/sysconfig/network

```
1  NETWORKING=yes
2  HOSTNAME=master1
```

Pliki konfiguracji sieciowych z katalogu /etc/sysconfig/network-scripts/

- ifcfg-eth0

```
1  DEVICE=eth0
2  BOOTPROTO=none
3  ONBOOT=yes
4  BRIDGE=hpc
```

- ifcfg-eth0.81

```
1  DEVICE=eth0.81
2  BOOTPROTO=none
3  ONBOOT=yes
4  BRIDGE=mgmt
5  VLAN=yes
```

- ifcfg-hpc

```
1  DEVICE=hpc
2  BOOTPROTO=none
3  ONBOOT=yes
4  TYPE=Bridge
5  IPADDR=192.168.4.11
6  NETMASK=255.255.252.0
7  #GATEWAY=192.168.4.1
```

- ifcfg-ib0

```
1  DEVICE=ib0
2  TYPE=InfiniBand
3  BOOTPROTO=none
4  ONBOOT=yes
5  IPADDR=192.168.12.11
```

```
NETMASK=255.255.252.0
MTU=65520
CONNECTED_MODE=yes
```

- ifcfg-ib1

```
DEVICE=ib1
BOOTPROTO=none
ONBOOT=no
```

- ifcfg-mgmt

```
DEVICE=mgmt
BOOTPROTO=none
ONBOOT=yes
TYPE=Bridge
IPADDR=172.31.71.249
NETMASK=255.255.255.0
GATEWAY=172.31.71.1
```

## 1.9 Serwer zarządzający – master2

### 1.9.1 Usługa networking

- /etc/sysconfig/network

```
NETWORKING=yes
HOSTNAME=master2
```

Pliki konfiguracji sieciowych z katalogu /etc/sysconfig/network-scripts/

- ifcfg-eth0

```
DEVICE=eth0
BOOTPROTO=none
ONBOOT=yes
BRIDGE=hpc
```

- ifcfg-eth0.81

```
DEVICE=eth0.81
BOOTPROTO=none
ONBOOT=yes
BRIDGE=mgmt
VLAN=yes
```

- ifcfg-hpc

```
DEVICE=hpc
BOOTPROTO=none
ONBOOT=yes
TYPE=Bridge
IPADDR=192.168.4.12
NETMASK=255.255.252.0
#GATEWAY=192.168.4.1
```

- ifcfg-ib0

```
DEVICE=ib0
TYPE=InfiniBand
BOOTPROTO=none
ONBOOT=yes
IPADDR=192.168.12.12
NETMASK=255.255.252.0
MTU=65520
CONNECTED_MODE=yes
```

- ifcfg-ib1

```
DEVICE=ib1
BOOTPROTO=none
ONBOOT=no
```

- ifcfg-mgmt

```
DEVICE=mgmt
BOOTPROTO=none
ONBOOT=yes
TYPE=Bridge
IPADDR=172.31.71.250
NETMASK=255.255.255.0
GATEWAY=172.31.71.1
```

## 1.10 Serwery zarządzające master – konfiguracja wspólna

### 1.10.1 Klaster HA – CMAN

- /etc/cluster/cluster.conf

```
<?xml version="1.0"?>
<cluster config_version="128" name="euros">
 <quorumd device="/dev/mapper/quorum" votes="6"/>
 <logging debug="off"/>
 <clusternodes>
      <clusternode name="master1" nodeid="1">
            <fence>
```

```
8                       <method name="ipmi">
9                               <device name="master1-m"/>
10                      </method>
11              </fence>
12      </clusternode>
13      <clusternode name="master2" nodeid="2">
14              <fence>
15                      <method name="ipmi">
16                              <device name="master2-m"/>
17                      </method>
18              </fence>
19      </clusternode>
20      <clusternode name="mds1" nodeid="3">
21              <fence>
22                      <method name="ipmi">
23                              <device name="mds1-m"/>
24                      </method>
25              </fence>
26      </clusternode>
27      <clusternode name="mds2" nodeid="4">
28              <fence>
29                      <method name="ipmi">
30                              <device name="mds2-m"/>
31                      </method>
32              </fence>
33      </clusternode>
34      <clusternode name="oss1" nodeid="5">
35              <fence>
36                      <method name="ipmi">
37                              <device name="oss1-m"/>
38                      </method>
39              </fence>
40      </clusternode>
41      <clusternode name="oss2" nodeid="6">
42              <fence>
43                      <method name="ipmi">
44                              <device name="oss2-m"/>
45                      </method>
46              </fence>
47      </clusternode>
48  </clusternodes>
49  <fencedevices>
50      <fencedevice agent="fence_ipmilan" auth="password" ipaddr="master1-m"
    lanplus="on" login="ipmi" name="master1-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
51      <fencedevice agent="fence_ipmilan" auth="password" ipaddr="master2-m"
    lanplus="on" login="ipmi" name="master2-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
52      <fencedevice agent="fence_ipmilan" auth="password" ipaddr="mds1-m"
    lanplus="on" login="ipmi" name="mds1-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
53      <fencedevice agent="fence_ipmilan" auth="password" ipaddr="mds2-m"
    lanplus="on" login="ipmi" name="mds2-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
54      <fencedevice agent="fence_ipmilan" auth="password" ipaddr="oss1-m"
    lanplus="on" login="ipmi" name="oss1-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
```

```
55        <fencedevice agent="fence_ipmilan" auth="password" ipaddr="oss2-m"
    lanplus="on" login="ipmi" name="oss2-m" passwd="fjio3pqX"
    privlvl="ADMINISTRATOR"/>
56    </fencedevices>
57    <rm>
58        <failoverdomains>
59              <failoverdomain name="master1" nofailback="0" ordered="1"
    restricted="1">
60                    <failoverdomainnode name="master1" priority="1"/>
61                    <failoverdomainnode name="master2" priority="2"/>
62              </failoverdomain>
63              <failoverdomain name="master2" nofailback="0" ordered="1"
    restricted="1">
64                    <failoverdomainnode name="master2" priority="1"/>
65                    <failoverdomainnode name="master1" priority="2"/>
66              </failoverdomain>
67              <failoverdomain name="mds1" nofailback="0" ordered="1"
    restricted="1">
68                    <failoverdomainnode name="mds1" priority="1"/>
69                    <failoverdomainnode name="mds2" priority="2"/>
70              </failoverdomain>
71              <failoverdomain name="mds2" nofailback="0" ordered="1"
    restricted="1">
72                    <failoverdomainnode name="mds2" priority="1"/>
73                    <failoverdomainnode name="mds1" priority="2"/>
74              </failoverdomain>
75              <failoverdomain name="oss1" nofailback="0" ordered="1"
    restricted="1">
76                    <failoverdomainnode name="oss1" priority="1"/>
77                    <failoverdomainnode name="oss2" priority="2"/>
78              </failoverdomain>
79              <failoverdomain name="oss2" nofailback="0" ordered="1"
    restricted="1">
80                    <failoverdomainnode name="oss2" priority="1"/>
81                    <failoverdomainnode name="oss1" priority="2"/>
82              </failoverdomain>
83        </failoverdomains>
84
85        <vm autostart="1" domain="master2" name="license"
    path="/etc/libvirt/qemu/"/>
86        <vm autostart="1" domain="master1" name="headnode"
    path="/etc/libvirt/qemu/"/>
87
88        <vm autostart="1" domain="master1" name="nethost"
    path="/etc/libvirt/qemu/"/>
89        <resources>
90              <lustrefs device="/dev/mapper/mdt" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-mdt" name="mdt" self_fence="1"/>
91              <lustrefs device="/dev/mapper/mgt" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-mgt" name="mgt" self_fence="1"/>
92              <lustrefs device="/dev/mapper/ost101" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-ost101" name="ost101"
    self_fence="1"/>
93              <lustrefs device="/dev/mapper/ost102" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-ost102" name="ost102"
    self_fence="1"/>
```

```
94              <lustrefs device="/dev/mapper/ost201" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-ost201" name="ost201"
    self_fence="1"/>
95              <lustrefs device="/dev/mapper/ost202" force_fsck="0"
    force_unmount="0" mountpoint="/mnt/lustre-ost202" name="ost202"
    self_fence="1"/>
96        </resources>
97        <service autostart="1" domain="mds1" exclusive="0" name="mgt"
    recovery="relocate">
98              <lustrefs ref="mgt"/>
99        </service>
100       <service autostart="1" domain="mds1" exclusive="0" name="mdt"
    recovery="relocate">
101             <lustrefs ref="mdt"/>
102       </service>
103       <service autostart="1" domain="oss1" exclusive="0" name="ost101"
    recovery="relocate">
104             <lustrefs ref="ost101"/>
105       </service>
106       <service autostart="1" domain="oss1" exclusive="0" name="ost102"
    recovery="relocate">
107             <lustrefs ref="ost102"/>
108       </service>
109       <service autostart="1" domain="oss2" exclusive="0" name="ost201"
    recovery="relocate">
110             <lustrefs ref="ost201"/>
111       </service>
112       <service autostart="1" domain="oss2" exclusive="0" name="ost202"
    recovery="relocate">
113             <lustrefs ref="ost202"/>
114       </service>
115  </rm>
116</cluster>
```

### 1.10.2 Usługa ntpd

- /etc/ntp.conf

```
117       driftfile /var/lib/ntp/drift
118       restrict default kod nomodify notrap nopeer noquery
119       restrict -6 default kod nomodify notrap nopeer noquery
120       restrict 127.0.0.1
121       restrict -6 ::1
122       server 10.91.91.13
123       server 10.91.91.14
124       server      127.127.1.0 # local clock
125       fudge 127.127.1.0 stratum 10
126       includefile /etc/ntp/crypto/pw
127       keys /etc/ntp/keys
```

- /etc/ntp/step-tickers

```
1  # List of servers used for initial synchronization.
```

```
2  10.91.91.13
3  10.91.91.14
```

## 1.11 Serwer dostępowy – headnode

### 1.11.1 Usługa networking

- /etc/sysconfig/network

```
1  NETWORKING=yes
2  HOSTNAME=headnode
```

Pliki konfiguracji sieciowych z katalogu /etc/sysconfig/network-scripts/

- ifcfg-eth0

```
1  DEVICE=eth0
2  BOOTPROTO=none
3  ONBOOT=yes
4  IPADDR=192.168.4.50
5  NETMASK=255.255.252.0
6  GATEWAY=192.168.4.1
```

### 1.11.2 Usługi pbs_server i pbs_sched (menadżer kolejkowania)

Konfiguracja serwera PBS częściowo przechowywana jest w wewnętrznej bazie danych aplikacji Torque. Z tego powodu nie wykonuje przed wykonaniem tradycyjnej kopii zapasowej plików konfiguracyjnych, należy posłużyć się poniższym poleceniem:

```
#  qmgr -c 'print server' > /etc/qmgr-backup.conf
```

Tak utworzony plik zawiera zdefiniowane kolejki oraz konfigurację serwera PBS. W celu przywrócenia konfiguracji należy wykonać polecenie:

```
#  qmgr < /etc/qmgr-backup.conf
```

- /etc/qmgr-backup.conf

```
1  #
2  # Create queues and set their attributes.
3  #
4  #
5  # Create and define queue batch
6  #
7  create queue batch
8  set queue batch queue_type = Execution
9  set queue batch resources_default.nodes = 1
10 set queue batch resources_default.walltime = 24:00:00
11 set queue batch enabled = True
```

```
set queue batch started = True
#
# Create and define queue fast
#
create queue fast
set queue fast queue_type = Execution
set queue fast resources_default.nodes = 1
set queue fast resources_default.walltime = 01:00:00
set queue fast enabled = True
set queue fast started = True
#
# Set server attributes.
#
set server scheduling = True
set server acl_hosts = headnode
set server acl_hosts += manager
set server acl_roots = root@*
set server managers = root@*
set server default_queue = batch
set server log_events = 511
set server mail_from = torque
set server query_other_jobs = True
set server scheduler_iteration = 600
set server node_ping_rate = 60
set server node_check_rate = 120
set server tcp_timeout = 300
set server job_stat_rate = 45
set server poll_jobs = True
set server mom_job_sync = True
set server keep_completed = 900
set server submit_hosts = headnode
set server submit_hosts += manager
set server allow_node_submit = True
set server log_keep_days = 30
set server next_job_number = 4452
set server record_job_info = True
set server record_job_script = True
set server job_log_keep_days = 30
set server moab_array_compatible = True
```

- /var/spool/torque/server_name

```
headnode
```

- /var/spool/torque/server_priv/nodes

```
n1 np=16 mpi
n2 np=16 mpi
n3 np=16 mpi
```

…

```
n95 np=16 mpi
```

```
4  n96 np=16 mpi
5  n97 np=16 mpi
```

- /var/spool/torque/sched_priv/sched_conf

```
3  round_robin: False      all
4  by_queue: True          prime
5  by_queue: True          non_prime
6  strict_fifo: false      ALL
7  fair_share: false ALL
8  help_starving_jobs      true  ALL
9  sort_queues true  ALL
10 load_balancing: false   ALL
11 sort_by: shortest_job_first   ALL
12 log_filter: 256
13 dedicated_prefix: ded
14 max_starve: 24:00:00
15 half_life: 24:00:00
16 unknown_shares: 10
17 sync_time: 1:00:00
```

#### 1.11.3 Usługa ssh

- /etc/ssh/sshd_config

```
1  AddressFamily inet
2  Protocol 2
3  SyslogFacility AUTHPRIV
4  HostbasedAuthentication yes
5  PasswordAuthentication yes
6  ChallengeResponseAuthentication yes
7  KerberosAuthentication yes
8  GSSAPIAuthentication yes
9  GSSAPICleanupCredentials yes
10 UsePAM yes
11 AcceptEnv LANG LC_CTYPE LC_NUMERIC LC_TIME LC_COLLATE LC_MONETARY
   LC_MESSAGES
12 AcceptEnv LC_PAPER LC_NAME LC_ADDRESS LC_TELEPHONE LC_MEASUREMENT
13 AcceptEnv LC_IDENTIFICATION LC_ALL LANGUAGE
14 AcceptEnv XMODIFIERS
15 X11Forwarding yes
16 Subsystem   sftp  /usr/libexec/openssh/sftp-server
```

#### 1.11.4 Usługa sssd (usługa katalogowa)

- /etc/sssd/sssd.conf

```
1  [sssd]
2  config_file_version = 2
3  services = nss, pam
4  domains = IMGW-ALADIN.LAN
5
```

```
[domain/IMGW-ALADIN.LAN]
ldap_id_use_start_tls = False
cache_credentials = True

enumerate = True
id_provider = ldap
auth_provider = krb5
chpass_provider = krb5
ldap_schema = rfc2307bis
ldap_force_upper_case_realm = True
ldap_user_object_class = user
ldap_group_object_class = group
ldap_user_gecos = displayName
ldap_user_home_directory = unixHomeDirectory

ldap_search_base = CN=USERS,DC=imgw-aladin,DC=lan
ldap_user_search_base =CN=USERS,DC=imgw-aladin,DC=lan
ldap_group_search_base =CN=GROUPS,DC=imgw-aladin,DC=lan
ldap_default_bind_dn = cn=bind,cn=Users,dc=imgw-aladin,dc=lan
ldap_default_authtok_type = password
ldap_default_authtok =
ldap_tls_cacertdir = /etc/openldap/cacerts


krb5_realm = IMGW-ALADIN.LAN
```

- /etc/nsswitch.conf

```
passwd:     files sss ldap
shadow:     files sss ldap
group:      files sss ldap
hosts:      files dns
bootparams: nisplus [NOTFOUND=return] files
ethers:     files
netmasks:   files
networks:   files
protocols:  files
rpc:        files
services:   files sss
netgroup:   files sss ldap
publickey:  nisplus
automount:  files ldap
aliases:    files nisplus
```

- /etc/openldap/ldap.conf

```
TLS_CACERTDIR /etc/openldap/cacerts

BASE CN=USERS,DC=imgw-aladin,DC=lan
```

- /etc/pam.d/system-auth-ac

```
1  auth        required      pam_env.so
2  auth        sufficient    pam_unix.so nullok try_first_pass
3  auth        requisite     pam_succeed_if.so uid >= 500 quiet
4  auth        sufficient    pam_sss.so use_first_pass
5  auth        required      pam_deny.so
6
7  account     required      pam_access.so
8  account     required      pam_unix.so broken_shadow
9  account     sufficient    pam_localuser.so
10 account     sufficient    pam_succeed_if.so uid < 500 quiet
11 account     [default=bad success=ok user_unknown=ignore] pam_sss.so
12 account     required      pam_permit.so
13
14 password    requisite     pam_cracklib.so try_first_pass retry=3
   type=
15 password    sufficient    pam_unix.so sha512 shadow nullok
   try_first_pass use_authtok
16 password    sufficient    pam_sss.so use_authtok
17 password    required      pam_deny.so
18
19 session     optional      pam_keyinit.so revoke
20 session     required      pam_limits.so
21 session     optional      pam_mkhomedir.so /etc/skel/ umask 0022
22 session     optional      pam_oddjob_mkhomedir.so
23 session     [success=1 default=ignore] pam_succeed_if.so service in
   crond quiet use_uid
24 session     required      pam_unix.so
25 session     optional      pam_sss.so
```

## 1.12 Serwer usług sieciowych – nethost

### 1.12.1 Usługa networking

- /etc/sysconfig/network

```
3  NETWORKING=yes
4  HOSTNAME=nethost
```

Pliki konfiguracji sieciowych z katalogu /etc/sysconfig/network-scripts/

- ifcfg-eth0

```
7  DEVICE=eth0
8  BOOTPROTO=none
9  ONBOOT=yes
10 IPADDR=192.168.4.50
11 NETMASK=255.255.252.0
12 GATEWAY=192.168.4.1
```

### 1.12.2 Usługi DNS, DHCP, TFTP (dnsmasq)

Oprogramowanie dnsmasq pracuje jako serwer DNS, DHCP i TFTP. Plik konfiguracyjny dnsmasq to /etc/dnsmasq.conf.d/hpc.conf. Usługa ta automatycznie tworzy rekordy DNS na podstawie zawartości pliku /etc/hosts oraz /etc/dhcp-hosts.conf.

dnsmasq świadczy usługę DNS dla adresów z poza klastra obliczeniowego za pośrednictwem serwerów zewnętrznych, które skonfigurowane są w /etc/dnsmasq.conf.d/hpc.conf i mają składnię pliku /etc/resolv.conf:

```
server=8.8.8.8
server=8.8.4.4
```

W obrębie klastra obliczeniowego (domena hpc lub hpc.local) dnsmasq świadczy usługę DNS orad DHCP na podstawie wpisów w pliku /etc/hosts i zawartości pliku /etc/dhcp-hosts.conf (mapowanie adresów mac na nazwy serwerów z /etc/hosts). Zawartość tego pliku ma format:

```
<adres mac>,<nazwa>
```

- /etc/dnsmasq.d/hpc.conf

```
1  local=/hpc/hpc.local/
2  interface=eth0
3  expand-hosts
4  domain=hpc.local
5  dhcp-range=192.168.7.1,192.168.7.254,12h
6  dhcp-no-override
7  no-resolv
8  server=/hpc/192.168.4.51
9  server=/hpc.local/192.168.4.51
10 server=8.8.8.8
11 server=8.8.4.4
12 dhcp-option=option:router,192.168.4.1
13 dhcp-authoritative
14 dhcp-option=144,n
15 enable-tftp
16 tftp-root=/tftpboot
17 dhcp-boot=pxelinux.0
18 dhcp-hostsfile=/etc/dhcp-hosts.conf
19 tftp-max=300
```

- /etc/hosts

```
127.0.0.1   localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.4.11   master1.hpc.local            master1.hpc           master1
192.168.4.12   master2.hpc.local            master2.hpc           master2
192.168.4.21   mds1.hpc.local        mds1.hpc              mds1
192.168.4.22   mds2.hpc.local        mds2.hpc              mds2
192.168.4.31   oss1.hpc.local        oss1.hpc              oss1
192.168.4.32   oss2.hpc.local        oss2.hpc              oss2
192.168.4.50   headnode.hpc.local           headnode.hpc          headnode
192.168.4.51   nethost.hpc.local            nethost.hpc           nethost
192.168.4.52   manager.hpc.local            manager.hpc           manager
192.168.4.53   license.hpc.local            license.hpc           license


192.168.12.11  master1-ib.hpc.local         master1-ib.hpc                master1-ib
192.168.12.12  master2-ib.hpc.local         master2-ib.hpc                master2-ib
192.168.12.21  mds1-ib.hpc.local            mds1-ib.hpc           mds1-ib
192.168.12.22  mds2-ib.hpc.local            mds2-ib.hpc           mds2-ib
192.168.12.31  oss1-ib.hpc.local            oss1-ib.hpc           oss1-ib
192.168.12.32  oss2-ib.hpc.local            oss2-ib.hpc           oss2-ib
172.31.71.131  master1-m.hpc.local          master1-m.hpc         master1-m
172.31.71.132  master2-m.hpc.local          master2-m.hpc         master2-m
172.31.71.133  mds1-m.hpc.local             mds1-m.hpc            mds1-m
172.31.71.134  mds2-m.hpc.local             mds2-m.hpc            mds2-m
172.31.71.135  oss1-m.hpc.local             oss1-m.hpc            oss1-m
172.31.71.136  oss2-m.hpc.local             oss2-m.hpc            oss2-m
172.31.71.249  master1-mgmt.hpc.local       master1-mgmt.hpc              master1-mgmt
172.31.71.250  master2-mgmt.hpc.local       master2-mgmt.hpc              master2-mgmt
172.31.71.251  mds1-mgmt.hpc.local          mds1-mgmt.hpc         mds1-mgmt
172.31.71.252  mds2-mgmt.hpc.local          mds2-mgmt.hpc         mds2-mgmt
172.31.71.253  oss1-mgmt.hpc.local          oss1-mgmt.hpc         oss1-mgmt
172.31.71.254  oss2-mgmt.hpc.local          oss2-mgmt.hpc         oss2-mgmt
192.168.5.1    n1.hpc.local          n1.hpc         n1
192.168.5.2    n2.hpc.local          n2.hpc         n2
192.168.5.3    n3.hpc.local          n3.hpc         n3
```

…

```
192.168.5.95   n95.hpc.local         n95.hpc        n95
192.168.5.96   n96.hpc.local         n96.hpc        n96
192.168.5.97   n97.hpc.local         n97.hpc        n97
192.168.13.1   n1-ib.hpc.local              n1-ib.hpc             n1-ib
192.168.13.2   n2-ib.hpc.local              n2-ib.hpc             n2-ib
192.168.13.3   n3-ib.hpc.local              n3-ib.hpc             n3-ib
```

…

```
192.168.13.95  n95-ib.hpc.local             n95-ib.hpc            n95-ib
192.168.13.96  n96-ib.hpc.local             n96-ib.hpc            n96-ib
192.168.13.97  n97-ib.hpc.local             n97-ib.hpc            n97-ib
172.31.71.141  n1-m.hpc.local        n1-m.hpc              n1-m
172.31.71.142  n2-m.hpc.local        n2-m.hpc              n2-m
172.31.71.143  n3-m.hpc.local        n3-m.hpc              n3-m
```

…

```
172.31.71.235  n95-m.hpc.local              n95-m.hpc             n95-m
172.31.71.236  n96-m.hpc.local              n96-m.hpc             n96-m
172.31.71.237  n97-m.hpc.local              n97-m.hpc             n97-m
```

- /etc/dhcp-hosts.conf

```
52:54:00:66:66:1d,ha1
52:54:00:9e:42:93,ha2
52:54:00:83:4f:51,license

D8:9D:67:73:13:C8,n1
D8:9D:67:73:59:B8,n2
D8:9D:67:73:00:A0,n3
```

…

```
D8:9D:67:73:B7:F8,n95
D8:9D:67:73:C0:58,n96
D8:9D:67:73:D9:08,n97

```

## 1.13 Serwer licencji – license

Ze względu na potrzebę uruchomienia usługi FLEXnet automatycznie, utworzony został skrypt lmgrd_intel. Podczas dodawania kolejnych usług lmgrd należy zweryfikować ich skrypty startowe pod kątem zgodności z istniejącymi.

### 1.13.1 Usługa lmgrd_intel

Usługa ta odpowiada za działanie serwera licencji sieciowych FLEXnet dla kompilatorów Intel.

```
#!/bin/sh
#
# lmgrd.intel FlexLM license manager for Intel compiler
#
# chkconfig: 345 99 1
# description: FlexLM license manager for Intel compiler
#

### BEGIN INIT INFO
# Provides: lmgrd.intel
# Required-Start: $network $local_fs $remote_fs
# Required-Stop: $network $local_fs
# Should-Start: $network $local_fs
# Should-Stop: $syslog
# Default-Start: 3 4 5
# Default-Stop: 0 1 2 6
# Short-Description: Start and stop FlexLM license manager for Intel compiler
# Description: Start and stop FlexLM license manager for Intel compiler
### END INIT INFO

# Source function library.
. /etc/init.d/functions

LMGRD_PATH="/opt/intel/flexlm"
LMUSER="flexlm"
LMGROUP="flexlm"

lmgrd="$LMGRD_PATH/lmgrd"
```

```
lmutil="$LMGRD_PATH/lmutil"
prog="lmgrd"
lockfile=/var/lock/subsys/$prog

LICENSE="$LMGRD_PATH/server.lic"
LOGFILE="/var/log/lmgrd.intel"

[ -e /etc/sysconfig/$prog ] && . /etc/sysconfig/$prog

lockfile=/var/lock/subsys/$prog

[ -x $lmgrd ] || exit 5
[ -x $lmutil ] || exit 5
[ -f $LICENSE ] || exit 6

checklog() {
   [ -f $LOGFILE ] || /bin/touch $LOGFILE
   /bin/chown $LMUSER:$LMGROUP $LOGFILE
}


start() {
echo -n $"Starting $prog: "
checklog
daemon --user $LMUSER $lmgrd -c $LICENSE -l $LOGFILE
retval=$?
echo
if [ $retval -eq 0 ] ; then
       touch $lockfile
fi
return $retval
}

stop() {
echo -n $"Stopping $prog: "
$lmutil lmdown -c $LICENSE -q >> $LOGFILE
retval=$?
echo
if [ $retval -eq 0 ] ; then
       rm -f $lockfile
       success
else
       failure
fi
return $retval
}

restart() {
stop
start
}

reload() {
restart
}

force_reload() {
restart
}

rh_status() {
$lmutil lmstat -a -c $LICENSE
return $?
}

```

```
94 rh_status_q() {
95 rh_status >/dev/null 2>&1
96 }
97
98 case "$1" in
99 start)
100 #               rh_status_q && exit 0
101         $1
102         ;;
103 stop)
104         rh_status_q || exit 0
105         $1
106         ;;
107 restart)
108         $1
109         ;;
110 reload)
111         rh_status_q || exit 7
112         $1
113         ;;
114 force-reload)
115         force_reload
116         ;;
117 status)
118         rh_status
119         ;;
120 restart)
121         $1
122         ;;
123 reload)
124         rh_status_q || exit 7
125         $1
126         ;;
127 force-reload)
128         force_reload
129         ;;
130 status)
131         rh_status
132         ;;
133 condrestart|try-restart)
134         rh_status_q || exit 0
135         restart
136         ;;
137 *)
138         echo $"Usage: $0 {start|stop|status|restart|condrestart|try-
    restart|reload|force-reload}"
139         exit 2
140 esac
141 exit $?
```

## 1.14 Węzły obliczeniowe

### 1.14.1 Usługa ssh

- /etc/ssh/shosts.equiv

```
1  +master1-ib.hpc.local
2  +master1.hpc.local
3  +master2-ib.hpc.local
4  +master2.hpc.local
5  +headnode.hpc.local
6  +n1-ib.hpc.local
```

```
+n1.hpc.local
+n2-ib.hpc.local
+n2.hpc.local
+n3-ib.hpc.local
+n3.hpc.local
```

…

```
+n95-ib.hpc.local
+n95.hpc.local
+n96-ib.hpc.local
+n96.hpc.local
+n97-ib.hpc.local
+n97.hpc.local
```

Plik ten należy rozszerzyć i rozpowszechnić na węzłach klastra przy każdorazowym dodawaniu węzłów obliczeniowych.

- /etc/ssh/sshd_config

```
AddressFamily inet
Protocol 2
SyslogFacility AUTHPRIV
HostbasedAuthentication yes
IgnoreUserKnownHosts yes
IgnoreRhosts yes
PasswordAuthentication no
ChallengeResponseAuthentication no
GSSAPIAuthentication no
GSSAPICleanupCredentials no
UsePAM yes
AcceptEnv LANG LC_CTYPE LC_NUMERIC LC_TIME LC_COLLATE LC_MONETARY LC_MESSAGES
AcceptEnv LC_PAPER LC_NAME LC_ADDRESS LC_TELEPHONE LC_MEASUREMENT
AcceptEnv LC_IDENTIFICATION LC_ALL LANGUAGE
AcceptEnv XMODIFIERS
X11Forwarding yes
Subsystem   sftp  /usr/libexec/openssh/sftp-server
```

- /etc/ssh/ssh_host_rsa_key

```
-----BEGIN RSA PRIVATE KEY-----
MIIEowIBAAKCAQEAv2xivZikInbLYIf4vqCoqC31huCSnyplcH/Rpr4DmTXFtXQa
5rlh96so5X1fdH2bKJmnT3cahaEaqGGFZVzd86CGA2fstEJ3iqtKy7HKu84d42wb
hM/cDaycs1Y5gb7wMY+gtAw4V8hTt1EyRV4oFTFKlkaQhBmLr9FlGaZgvJ082Wx2
jRUqYuJ9mqRTZZYNjTd1++cU1lwEdqJwlTFRHEyz/ZdOo68F/KeYB3tLIzPG9F+A
nEf2lF1YM9b3OUmECjVr5cklBTZm0eLfdSu1gjCkIveFgj2tMSIu+U2ddLXzGLiB
2Qz8KVwfNARQJi4do+Lp+T/Uk4bSFBYuY00eOwIBIwKCAQEAg0MB4Rg2CQEAfLUC
VtSQ6FoHcm4bZdPtyXw3/U8YaQ7t+MSku3fOJiw5TOhBdHNjFIaenOP07erQc3YS
VCJsT0mAd100mN0exXzFkv2SVOUbzyWPNn/nWdVyxB3eSlcLGquhZYS467yRLT79
5mxzQbtmWGrmwPuS/DfQTBpQ9lv+lolQ/nQKxayho9HUaKf5OmnRIPivDIVvYGNu
```

```
8mM8pXFsA8XqTeZ/e+/pefrh6Tr9mx3w9013zglBiBq3sqYjAkDkIMaz8sNO2IUx
mhGDlcEAkJXSb1WerdIldVP3/oqWgD55mwIYFFe4rXjZ26uTJO19HOBkYiq9cfcE
g+/mCwKBgQD96E+5q3qG1V2OWXka/1vRfosbK2LNJUgISrJaDJaqe5GRof0//1Zb
LnnF/TGYeIYNyOm3axv/g9rFnBoJzAbBGoW/DsH4hmoEa9oKY6GjCxcBnzffDDpj
aPrla4dLpgWDDapjwkNqcZMGZyRHqqzOgIb5b/FkjyHD2nMwG7LmaQKBgQDBAD8m
w9wT1+kuMkPiou/wmXy0gG7zFCa5v18UvHSY2iBk1hluSBMw5BqdjbvT4Tex/pTJ
yIWoWV+9sUChiLslNxEK0eVmEho6FPFCENsykgrXBjp+PlEXtVTS7VaIJiZUkAhX
ahEumh6Az+UgGycYjYGaFIcy9TwPSMVtBzGzAwKBgQDvZgIG1OE9TNSNh49xOejb
d0/1C6ZMZPrF/UkTEy74gyLZxJ5SSISfHTD8gQLgNx9O02AEtXIrbbEDdeyqJsSK
LvMh2rbjAmP1ikKcFMuZuf+/s2CQeT5dusCsbLLDqynEtRz++P28IfEGCXn6bbjf
9YaTaYvT12j6fYKMceMxAwKBgEIr+GUP868IMrDAxtFcb4Ww945X64aKkO86THwj
Wy0XlhPxolGyT7j9v/uBDTKznhEkFcGGlDm4PhUm1FSj6GSH6JX+0k7hopeSJtTS
lEvaTNwCIq7wy1iV8TJgAGkxpr3oPV/MmC08J7citPUQn7Cli4VI3eWV6LTIfjQC
d3CTAoGBAJVsXA1eElTcDH2/eECUGHj5Mivq7o0OnajiBWeOauH4Uy46r1Sm1SY6
L8uitGO0PWwZ07M0T4RD4sRTh0VMeHthFSsm5pJqrMowNjOIdpADjkD1gmuD/kEm
p/eCsqkGI9xgFPN6fnQG7cGZqkVPdyYYQcI8JbTis7NTvwvBcB2s
-----END RSA PRIVATE KEY-----

```

- /etc/ssh/ssh_host_rsa_key.pub

```
ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEAv2xivZikInbLYIf4vqCoqC31huCSnyplcH/Rpr4Dm
TXFtXQa5rlh96so5X1fdH2bKJmnT3cahaEaqGGFZVzd86CGA2fstEJ3iqtKy7HKu84d42
wbhM/cDaycs1Y5gb7wMY+gtAw4V8hTt1EyRV4oFTFKlkaQhBmLr9FlGaZgvJ082Wx2jRU
qYuJ9mqRTZZYNjTd1++cU1lwEdqJwlTFRHEyz/ZdOo68F/KeYB3tLIzPG9F+AnEf2lF1Y
M9b3OUmECjVr5cklBTZm0eLfdSu1gjCkIveFgj2tMSIu+U2ddLXzGLiB2Qz8KVwfNARQJ
i4do+Lp+T/Uk4bSFBYuY00eOw==
```

- /etc/ssh/ssh_known_hosts2

```
n*,192.168.5.*,192.168.13.*                                    ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEAv2xivZikInbLYIf4vqCoqC31huCSnyplcH/Rpr4Dm
TXFtXQa5rlh96so5X1fdH2bKJmnT3cahaEaqGGFZVzd86CGA2fstEJ3iqtKy7HKu84d42
wbhM/cDaycs1Y5gb7wMY+gtAw4V8hTt1EyRV4oFTFKlkaQhBmLr9FlGaZgvJ082Wx2jRU
qYuJ9mqRTZZYNjTd1++cU1lwEdqJwlTFRHEyz/ZdOo68F/KeYB3tLIzPG9F+AnEf2lF1Y
M9b3OUmECjVr5cklBTZm0eLfdSu1gjCkIveFgj2tMSIu+U2ddLXzGLiB2Qz8KVwfNARQJ
i4do+Lp+T/Uk4bSFBYuY00eOw==
headnode*,192.168.4.50                                         ssh-rsa
AAAAB3NzaC1yc2EAAAABIwAAAQEArPDP1wQhjAjeBzm7QPOsSmTqKW//1P4nMD+AJSdy7
0TI0k27hY2QB316XfzTpSR3K0blWN0R6LwvJ2bYG38BWKoS2KquxA9vgdXTNPcejRc6nv
xQgTtR/uow72ZSwAyocloQvo/o2K9tBbyhaOQsCRJKdTHHxjfowGADtYh/12LyeaZDwFi
k5o1LdUBN84bMsETSF9xGcXpP77RQLFjCyRglGPLaAvHYPj4ATfkN90RAF8wb/EHWqa4l
IBQg3JoSyFH6PLQLqyjH5jPb0eAfSZsxhVJ7KM7LYRbDcjwvKVTCv6NIiaoJIMn7f6WCw
okFwL9meWPBhmSvNV1Kl5c9EQ==
```

### 1.14.2 Usługa sssd (usługa katalogowa)

- /etc/sssd/sssd.conf

```
[sssd]
config_file_version = 2
```

```
services = nss, pam
domains = IMGW-ALADIN.LAN
[nss]
[pam]
[domain/IMGW-ALADIN.LAN]
ldap_id_use_start_tls = False
cache_credentials = True
enumerate = True
id_provider = ldap
auth_provider = krb5
chpass_provider = krb5
ldap_schema = rfc2307bis
ldap_force_upper_case_realm = True
ldap_user_object_class = user
ldap_group_object_class = group
ldap_user_gecos = displayName
ldap_user_home_directory = unixHomeDirectory

ldap_search_base = CN=USERS,DC=imgw-aladin,DC=lan
ldap_user_search_base =CN=USERS,DC=imgw-aladin,DC=lan
ldap_group_search_base =CN=GROUPS,DC=imgw-aladin,DC=lan
ldap_default_bind_dn = cn=bind,cn=Users,dc=imgw-aladin,dc=lan
ldap_default_authtok_type = password
ldap_default_authtok =
ldap_tls_cacertdir = /etc/openldap/cacerts


krb5_realm = IMGW-ALADIN.LAN
```

- /etc/nsswitch.conf

```
passwd:     files sss ldap
shadow:     files sss ldap
group:      files sss ldap
hosts:      files dns
bootparams: nisplus [NOTFOUND=return] files
ethers:     files
netmasks:   files
networks:   files
protocols:  files
rpc:        files
services:   files sss
netgroup:   files sss ldap
publickey:  nisplus
automount:  files ldap
aliases:    files nisplus
```

- /etc/openldap/ldap.conf

```
TLS_CACERTDIR /etc/openldap/cacerts

BASE CN=USERS,DC=imgw-aladin,DC=lan
```

- /etc/pam.d/system-auth-ac

```
1  auth        required      pam_env.so
2  auth        sufficient    pam_fprintd.so
3  auth        sufficient    pam_unix.so nullok try_first_pass
4  auth        sufficient    pam_sss.so use_first_pass
5  auth        sufficient    pam_krb5.so use_first_pass
6  auth        sufficient    pam_ldap.so use_first_pass
7  auth        sufficient    pam_winbind.so use_first_pass
8  auth        required      pam_deny.so
9  account     required      pam_access.so
10 account     required      pam_unix.so broken_shadow
11 account     sufficient    pam_localuser.so
12 account     sufficient    pam_succeed_if.so uid < 500 quiet
13 account     [default=bad success=ok user_unknown=ignore] pam_sss.so
14 account     [default=bad success=ok user_unknown=ignore] pam_ldap.so
15 account     [default=bad success=ok user_unknown=ignore] pam_krb5.so
16 account     [default=bad success=ok user_unknown=ignore]
   pam_winbind.so
17 account     required      pam_permit.so
18 password    requisite     pam_cracklib.so try_first_pass retry=3
   type=
19 password    sufficient    pam_unix.so sha512 shadow nis nullok
   try_first_pass use_authtok
20 password    sufficient    pam_sss.so use_authtok
21 password    sufficient    pam_krb5.so use_authtok
22 password    sufficient    pam_ldap.so use_authtok
23 password    sufficient    pam_winbind.so use_authtok
24 password    required      pam_deny.so
25 session     optional      pam_keyinit.so revoke
26 session     required      pam_limits.so
27 session     optional      pam_oddjob_mkhomedir.so
28 session     [success=1 default=ignore] pam_succeed_if.so service in
   crond quiet use_uid
29 session     required      pam_unix.so
30 session     optional      pam_sss.so
31 session     optional      pam_krb5.so
session     optional      pam_ldap.so
```

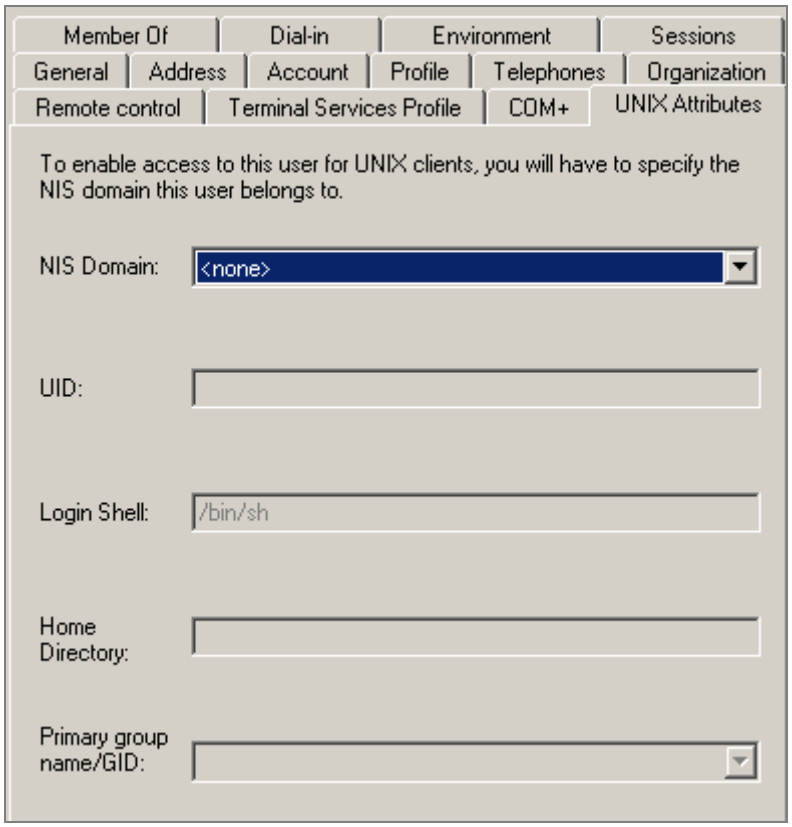
# 8. Procedury eksploatacyjne

W poniższych punktach opisano standardowe procedury umożliwiające weryfikację pracy poszczególnych elementów systemu.

## 1.15 Zarządzanie użytkownikami

### 1.15.1 Dodawanie użytkowników

Dodawanie użytkowników odbywa się z poziomu kontrolera domeny Windows, przypisując odpowiednie atrybuty dla istniejących kont w domenie IMGW.




Aby umożliwić użytkownikowi należy we właściwościach użytkownika uzupełnić kartę `Unix Attributes`, według następującego schematu.

- W polu `NIS Domain` należy wybrać domenę do której dodawany będzie użytkownik
- W polu `UID` należy podać pierwszy wolny identyfikator użytkownika
- W polu `Login Shell` należy podać preferowaną powłokę systemową użytkownika, zalecaną powłoką jest `/bin/bash`
- W polu Home Directory należy podać ścieżkę do katalogu domowego użytkownika
- W polu `Primary group name/GID` należy wybrać z listy grupę `hpc` (id=5000)

Dla prawidłowej pracy klastra wymagane jest by katalogi domowe użytkowników znajdowały się w folderze `/home` zlokalizowanym na współdzielonym zasobie sieciowym.

## 1.16 Zarządzanie klastrem

### 1.16.1 Sposób zarządzania

Węzły zostały wyposażone w moduł zarządzający (BMC) zgodny ze standardem IPMI 2.0. Umożliwia to zarządzanie węzłami za pomocą narzędzi dostępnych w systemie operacyjnym.

Na każdym serwerze wyposażonym w moduł BMC utworzony został użytkownik **ipmi**.

### 1.16.2 Adresacja sieciowa modułów zarządzających

Dostęp do zarządzania modułami BMC możliwy jest jedynie z serwerów zarządzających `master`.

Adresacja sieciowa modułów opisana jest w Tabela 2. Adresacja zarządzania fizycznego w punkcie 3.3.

### 1.16.3 Wydawanie poleceń zarządzających

Aby ułatwić procedurę wydawania poleceń modułów zarządzających przygotowano skrypt `ipmi`, którego celem jest ułatwienie korzystania z narzędzia `ipmitool`. Za pomocą skryptu `ipmi` można wydawać polecenia modułom zarządzającym bez konieczności każdorazowej autoryzacji oraz znajomości adresów kart zarządzających.

Składnia polecenia ma postać:

```
# ipmi <nazwa_węzła> <polecenie_ipmitool>
```

Przykładowe wywołanie - wyłączenie węzła n3

```
# ipmi n3 power off
```

Aby uzyskać więcej możliwych poleceń należy wywołać skrypt bez podawania polecenia. Przykładowe polecenia `ipmi`:

- `power`
  - `on` – włączenie
  - `off` – wyłączenie
  - `reset` – reset
  - `cycle` – ponowne uruchomienie poprzez całkowite wyłączenie, a następnie włączenie serwera
  - `soft` – wyłączenie z wcześniejszym zamknięciem systemu
- `sol activate` – uruchomienie przekierowania konsoli zdalnej, wyjście z konsoli następuje za pomocą sekwencji klawiszy `~.` poprzedzonych klawiszem `Enter` (pełna lista komend dostępna jest po wprowadzeniu klawiszy `~?` poprzedzonych klawiszem `Enter`).

- `sel list` – odczyt logu sprzętowego
- `sdr` – odczyt czujników systemowych

### 1.16.4 Wydawanie poleceń dla wszystkich węzłów

Aby wydać polecenia dla wszystkich węzłów w klastrze należy użyć powłoki pdsh. Wywołanie powłoki ma postać:

```
#  pdsh -w n[1-96] <polecenie_ssh>
```

odpowiednio dla węzłów n1, n{…} ,n96

Wywołanie polecenia ipmi dla całego klastra ma postać:

```
#  pdsh -w n[1-96] -R exec ipmi %h <polecenie_ipmi>
```

odpowiednio dla węzłów n1, n{…} ,n96.

Nie można zarządzać BMC z poziomu węzłów obliczeniowych

### 1.16.5 Procedura wyłączenia oraz ponownego włączenia środowiska

Prawidłowe wyłączenie środowiska składa się z trzech etapów:

**a) Wyłączenie węzłów obliczeniowych**

Wyłączenie węzłów obliczeniowych odbywa się wydając komendę z serwera `master1` lub `nethost`:

```
#  pdsh -w n[1-96] poweroff
```

lub korzystając ze sterowania modułami zarządzającymi

```
#  pdsh -w n[1-96] -R exec ipmi %h power soft
```

Przed wykonaniem kolejnego kroku należy zweryfikować poprawność wyłączenia wszystkich węzłów obliczeniowych, poleceniem:

```
#  pdsh -w n[1-96] -R exec ipmi %h power status
```

Wynik polecenia powinien zawierać informację „Chassis Power is off” dla każdego z węzłów.

**b) Wyłączenie maszyn wirtualnych**

Polecenia te służą jedynie do wyłączenia maszyn wirtualnych działających poza klastrem wysokiej dostępności

Wyłączenie maszyn wirtualnych należy wykonać wydając z serwerów master poleceniem:

```
# virsh shutdown <nazwa_vm>
```

dla każdej maszyny wirtualnej widocznej w wyniku polecenia:

```
# virsh list
```

Alternatywnie można posłużyć się skryptem `virt-shutdown-all`, wykrywającym uruchomione maszyny wirtualne i wyłączającym je.

Zawartość skryptu `virt-shutdown-all`:

```
3  #!/bin/bash
4
5  for vm in `virsh list | sed 1,2d | awk {' print $2 '}`; do
6  virsh shutdown $vm
7  done
```

Przed wykonaniem ostatniego kroku należy upewnić się, że wszystkie maszyny wirtualne zostały wyłączone, wydając ponownie polecenie:

```
# virsh list --all
```

Wszystkie pozycje na wyświetlonej liście powinny mieć stan „shut off”.

W przypadku gdy maszyna wirtualna nie reaguje na komendę wyłączenia, należy przeprowadzić jej ręczne zamknięcie korzystając z aplikacji `virt-manager` lub logując się zdalnie do danego systemu.

**c) Zatrzymanie klastra HA**

Z dowolnego serwera znajdującego się w klastrze HA należy wydać polecenie:

```
# ccs –f /etc/cluster/cluster.conf --stopall
```

**d) Wyłączenie serwerów zarządzających**

Serwery zarządzające należy wyłączyć logując się na każdy poprzez SSH i wydając polecenie:

```
# poweroff
```

Prawidłowe wyłączenie serwerów można zweryfikować z dowolnego komputera mającego dostęp do sieci zarządzającej, korzystając z poleceń IPMI lub logując się na stronę internetową pod adresem IP zarządzania danego serwera.

Wyłączenie serwerów master bez wcześniejszego wyłączenia maszyn wirtualnych spowoduje zapisanie stanu maszyn wirtualnych

## 1.17 Maszyny wirtualne

Konfiguracja maszyn wirtualnych znajduje się w katalogu `/etc/libvirt/qemu`. Zarządzanie maszynami jest zależne od tego czy dana maszyna pracuje pod kontrolą klastra HA czy poza nim.

### 1.17.1 Działające poza klastrem wysokiej dostępności

Do zarządzania maszynami wirtualnymi służą programy graficzne `virt-manager` oraz `virt-viewer`, a także konsola tekstowa `virsh`. Przydatne polecenia konsoli tekstowej:

- `virsh console <nazwa domeny>` – podłączenie do portu szeregowego maszyny wirtualnej, wyjście z konsoli następuje za pomocą kombinacji klawiszy `<CTRL> plus ]`
- `virsh start <nazwa domeny>` – uruchamia maszynę wirtualną
- `virsh shutdown <nazwa domeny>` – zatrzymuje maszynę wirtualną (poprzez wysłanie sygnału do systemu operacyjnego maszyny)
- `virsh destroy <nazwa domeny>` – zatrzymuje maszynę wirtualną (bez oczekiwania na wyłączenie systemu operacyjnego) – użycie tego polecenia w czasie normalnej pracy jest niezalecane
- `virsh create <ścieżka do pliku xml>` – tworzy nową maszynę wirtualną lub uaktualnia konfigurację aktualnej maszyny na podstawie zawartości pliku.
- `virsh list` – wyświetla listę maszyn wirtualnych,
  - `--active`, tylko aktywne (domyślny)
  - `--inactive`, tylko nieaktywne
  - `--all`, aktywne oraz nieaktywne

### 1.17.2 Działające w klastrze wysokiej dostępności

**a) Stan maszyn wirtualnych**

Ogólne informacje o stanie maszyn wirtualnych dostępne są po wykonaniu polecenia:

```
# clustat
```

Bardziej szczegółowy raport, zawierający czas ostatniej zmiany stanu domeny, aktualnego oraz poprzedniego właściciela usługi oraz dodatkowe flagi, dostępne są poprzez polecenie:

```
# clustat -l
```

**b) Migracja maszyn wirtualnych**

Migracja maszyny na konkretny host znajdujący się w klastrze HA:

```
# clusvcadm –M vm:<nazwa_wm> -m <host>
```

Przykład:

```
# clusvcadm -M vm:license -m master2
```

### c) Relokacja maszyn wirtualnych

Relokacja powoduje przeniesienie maszyny na innego gospodarza w ramach tej samej domeny:

```
# clusvcadm -r vm:<nazwa_wm>
```

### d) Dodawanie maszyn wirtualnych do klastra HA

Podstawowe wymagane informacje dotyczące dodawania maszyny wirtualnej to:

```
# ccs -h <host> --addvm <nazwa_vm> path=<ścieżka_do_xml>
```

Przykład:

```
# ccs -h master1 --addvm guest1 path=/etc/libvirt/qemu
```

Dodatkowo można zdefiniować następujące parametry:

| Nazwa parametru | Opis | Akceptowane wartości |
|---|---|---|
| name | Nazwa maszyny wirtualnej | string |
| domain | Nazwa domeny failover | string |
| autostart | Automatyczne uruchamianie domeny po osiągnięciu quorum | bool (domyślnie 1) |
| exclusive | Relokuje maszynę tylko na węzeł na którym nie są uruchomione inne zasoby | bool (domyślnie 0) |
| recovery | Polityka odzyskiwania maszyny wirtualnej:<br>• restart – uruchamia ponownie maszynę na hoście zanim spróbuję przenieść go na inny<br>• relocate – przenosi maszynę na inny host niezwłocznie po awarii<br>• disable – nie odzyskuje maszyny po awarii | string (domyślnie restart) |
| use_virsh | Korzystanie z virsh w miejscu xm dla hypervisor'a Xen (zbędne dla KVM) | integer (domyślnie 0) |
| migrate | Rodzaj stosowanej migracji<br>• live – migracja na żywo<br>• paused – migracja offline | string (domyślnie live) |
| tunnelled | Tunelowanie migracji przez SSH w celu bezpiecznej migracji | string |

| | | |
|---|---|---|
| path | Ścieżka do plików konfiguracyjnych xml | string |
| snapshot | Ścieżka dla migawek | string |
| depend_mode | Tryb zależności<br>• hard – usługa jest zatrzymywana jeśli jej zależność jest zatrzymana<br>• soft – usługa wymaga zależności tylko podczas uruchoamiania | String (domyślnie hard) |
| max_restarts | Maksymalna ilość restartów dla danej usługi | string (domyślnie 0) |
| restart_expire_time | Czas po jakim informacja o wcześniejszych restartach wygasa | string (domyślnie 0) |
| status_program | Dodatkowe sprawdzenie stanu aplikacji wewnątrz maszyny wirtualnej | string |
| hypervisor | Dodatkowe parametry dla hypervisora | string (domyślnie auto) |
| hypervisor_uri | Adres URI hypervisora | string (domyślnie auto) |
| migration_uri | Adres URI używany podczas migracji | string (domyślnie auto) |

**e) Aktywowanie maszyn wirtualnych w klastrze HA**

Aktywowanie nieaktywnej usługi:

```
# clusvcadm -e vm:<nazwa_wm>
```

opcjonalnie można podać docelowy host:

```
# clusvcadm -e vm:<nazwa_wm> -m <host>
```

**f) Deaktywowanie maszyn wirtualnych w klastrze HA**

Deaktywowanie aktywnej usługi:

```
# clusvcadm -d vm:<nazwa_wm>
```

**g) Zatrzymanie maszyn wirtualnych w klastrze HA**

```
# clusvcadm -s vm:<nazwa_wm>
```

**h) Zamrożenie monitorowania maszyn wirtualnych w klastrze HA**

W celu operacji administracyjnych istnieje możliwość tymczasowe zawieszania monitorowania maszyny wirtualnej. Maszyna wirtualna podczas tej operacji nie zostaje zatrzymana, a jedynie przestaje być śledzona przez klaster.

```
# clusvcadm -Z vm:<nazwa_wm>
```

**i) Odmrożenie maszyn wirtualnych w klastrze HA**

Zamrożoną maszynę należy odmrozić po zakończeniu wymaganych operacji administracyjnych:

```
# clusvcadm -U vm:<nazwa_wm>
```

**j) Ponowne uruchomienie maszyn wirtualnych w klastrze HA**

W miarę możliwości maszynę wirtualną należy restartować korzystając z mechanizmów systemowych lub polecenia virsh. Jeśli wymagane jest ponowne uruchomienie usługi w klastrze HA, wykonuje się je następującym poleceniem:

```
# clusvcadm -R vm:<nazwa_wm>
```

**k) zmiana konfiguracji klastra HA**

Plik konfiguracyjny klastra HA znajduje się na każdym węźle zarządzającym. Aby zmienić konfigurację należy wymedytować pliki /etc/cluster/cluster.conf. Po dokonaniu modyfikacji należy zsynchronizować plik konfiguracyjny pomiędzy serwerami poleceniem

```
# ccs_sync –i
```

a następnie zaktualizować aktywną konfigurację poleceniem:

```
# cman_tool version -r
```

## 1.18 Instalacja i aktualizacja oprogramowania

Instalacja oprogramowania na węzłach zarządzających odbywa się korzystając ze standardowych mechanizmów systemu operacyjnego – zalecane jest korzystanie z polecania `yum` oraz `rpm`.

Dla węzłów obliczeniowych należy wykorzystać mechanizm `chroot`, który można zastosować z węzłów zarządzających `master` lub węzła dostępowego `headnode`.

Przykładowe uruchomienie środowiska chroot:

```
# mount -t proc proc /mnt/lustre/diskless/sl64-1/proc/
# mount -t sysfs sys /mnt/lustre/diskless/sl64-1/sys/
# mount -o bind /dev /mnt/lustre/diskless/sl64-1/dev
```

```
# chroot /mnt/lustre/diskless/sl64-1/
```

W tak przygotowanym środowisku można instalować oprogramowanie zgodnie ze standardową procedurą systemu operacyjnego.

W celu aktualizacji klienta Lustre należy w pierwszej kolejności zbudować odpowiednią paczkę w poniższy sposób.

Na przykładzie Lustre 2.4.2 na CentOS 6.5 z kernel 2.6.32-431.5.1.el6.x86_64:

```
# wget -P ~/rpmbuild/SRPMS
  http://downloads.whamcloud.com/public/lustre/latest-
  maintenance-release/el6/client/SRPMS/lustre-client-2.4.2-
  2.6.32_358.23.2.el6.x86_64.src.rpm
# rpmbuild --define 'configure_args --disable-server' --define
  'lustre_name lustre-client' --define 'kversion 2.6.32-
  431.5.1.el6.x86_64' --define 'kdir /usr/src/kernels/2.6.32-
  431.5.1.el6.x86_64/' --rebuild ~/rpmbuild/SRPMS/lustre-client-
  2.4.2-2.6.32_358.23.2.el6.x86_64.src.rpm
```

Wynikowe rpm'y znajdują się w katalogu ~/rpmbuild/RPMS

## 1.19 Dodawanie nowego węzła do puli obliczeniowej

### 1.19.1 Konfiguracja BIOS

Podczas instalacji nowego węzła obliczeniowego, należy zaktualizować firmware do najnowszej wersji oraz, jeśli to możliwe, wyłączyć HyperThreading i ustawić poniższą kolejność urządzeń rozruchu:

1) Network boot (PXE)

Moduł zarządzający BMC, jeśli istnieje, powinien być skonfigurowany na podstawie domyślnych ustawień, jedynie definiując statyczną konfigurację sieci:

- adres IP w puli 172.31.71.131/25
- maska sieciowa 255.255.255.0
- brama 172.31.71.1

### 1.19.2 Dodanie węzła do puli obliczeniowej

Następnie na maszynie wirtualnej headnode należy ręcznie dodać wpis dotyczący węzła z pliku `/var/spool/torque/server_priv/nodes` lub wykonać następujące polecenia:

```
# cd /var/spool/torque/server_priv/
# echo "<nazwa_wezla> np=<x> <y>" >> nodes
```

Gdzie <x> to suma liczby rdzeni procesorów znajdujących się w węźle, a <y> dostępnymi parametrami (np. `mpi` lub gpu).

Następnie należy odświeżyć konfigurację usługi pbs_server:

```
# service pbs_server restart
```

Weryfikację poprawności działania konfiguracji należy wykonać poniższym poleceniem:

```
# pbsnodes <nazwa_wezla>
```

Wynik polecenia powinien zawierać linię „state = free”.

W przypadku gdy nowododany węzeł nie zmienia stanu z „down” na „free” należy zweryfikować czy usługi zapory sieciowej iptables oraz ip6tables są wyłączone.

#### 1.19.3 Aktualizacja DNS

W celu aktualizacji usługi DNS należy na dowolnym serwerze zarządzającym master należy przejść do katalogu /root/scripts/hosts i wyedytować plik hosts.config. Następnie należy uruchomić skrypty:

```
# ./generate_hosts_file.sh
# ./distribute_hosts_file.sh
```

W celu wprowadzenia zmian należy zalogować się na maszynę nethost i wykonać restart usługi dnsmasq poleceniem:

```
# service dnsmasq restart
```

#### 1.19.4 Aktualizacja DHCP

W celu dodania lub zmian adresów MAC należy zalogować się do serwera nethost i wyedytować plik /etc/dhcp-hosts.conf., wpisując MAC adres i krótką nazwę domenową hosta w formacie:

```
D8:9D:67:73:13:C8,n1
```

Gdzie D8:9D:67:73:13:C8 to adres MAC, a n1 to nazwa domenowa węzła.

### 1.20 Usuwanie węzła obliczeniowego z puli obliczeniowej

Przed usunięciem węzła z puli obliczeniowej należy ustawić jego stan na „offline” w menadżerze kolejkowania. W tym celu należy wydać polecenie:

```
# pbsnodes -o <nazwa_węzła>
```

Następnie na maszynie wirtualnej headnode należy ręcznie usunąć wpis dotyczący węzła z pliku /var/spool/torque/server_priv/nodes lub posłużyć się poniższym poleceniem:

```
# sed -i /<nazwa_wezla>/d /var/spool/torque/server_priv/nodes
```

Następnie należy ponownie uruchomić usługę `pbs_server`:

```
# service pbs_server restart
```

## 1.21 Moduły środowiskowe

Oprogramowanie Environmental Modules udostępnia prosty mechanizm zarządzania zmiennymi środowiskowymi przez użytkownika.

### 1.21.1 Wyświetlanie listy dostępnych modułów

Wszystkie dostępne moduły wraz ze ścieżkami lokalizacji dostępne są po wydaniu polecenia:

```
# module avail
```

Moduły dostępne na klastrze:

```
------------- /usr/share/Modules/modulefiles --------------
dot         module-info null
module-cvs  modules     use.own

----- /usr/share/Modules/modulefiles_euros/compilers ------
PE-gcc              intel/13.0.1(default)
PE-intel

-------- /usr/share/Modules/modulefiles_euros/mpi ---------
intel-mpi/4.1.0            openmpi/1.6.4-intel
openmpi/1.6.4-gcc(default)
```

### 1.21.2 Dodawanie nowych modułów

Tworzenie nowych modułów sprowadza się do utworzenia pliku w jednym ze zdefiniowanych katalogów, będącego skryptem w języku TCL. Poniżej znajduje się przykład modułu udostępniającego użytkownikowi Toolkit oraz SDK technologii CUDA:

```
#%Module1.0

proc ModulesHelp { } {
  puts stderr "NVIDIA CUDA Toolkit & SDK"
  puts stderr "TAGS: nvidia,cuda"
}

module-whatis "NVIDIA CUDA Toolkit & SDK"

prepend-path PATH               /usr/local/cuda-5.0/bin
prepend-path LD_LIBRARY_PATH    /usr/local/cuda-5.0/lib64
prepend-path INCLUDE         /usr/local/cuda-5.0/include
```

Tak utworzony moduł spowoduje tymczasową modyfikację zmiennej środowiskowej PATH, LD_LIBRARY_PATH oraz INCLUDE.

Procedura ModulesHelp wyświetla informacje o module w przypadku użycia komend:

```
# module whatis cuda
```

lub

```
# module help cuda
```

### 1.21.3 Definiowanie domyślnej wersji

W celu ułatwienia ładowania modułów dostępnych w więcej niż jednej wersji, możliwe jest ustalenie domyślnie ładowanego modułu w przypadku gdy użytkownik nie określi konkretnej wersji. Tak zdefiniowane moduły posiadają automatycznie generowany dopisek (default).

W celu określenia domyślnej wersji modułu należy w katalogu zawierającym definicję modułu utworzyć plik o nazwie ".version", o treści:

```
23 #%Module1.0
9  set     ModulesVersion  "nazwa_pliku_modułu"
```

Przykład definiujący moduł PGI/10.9 jako domyślny:

- /usr/share/Modules/modulefiles_euros/compilers/pgi/.version

```
24 #%Module1.0
10 set     ModulesVersion  "10.9"
```

## 1.22 Monitorowanie dzienników zdarzeń

Dla każdej z maszyn fizycznych najważniejsze logi systemowe przesyłane są na zdalny serwer wirtualny – loghost. Gromadzone są one w katalogu /var/log/remote/nazwa_hosta.

Pozostałe logi dostępne są na odpowiednich węzłach ze względu na potrzebę ich odczytywania przez dostarczone narzędzia.

### 1.22.1 Zadania menadżera kolejkowania

Informacje dotyczące uruchomionych oraz niedawno zakończonych zadań dostępne są z węzła headnode:

```
# qstat –an
```

Sprawdzenie informacji o konkretnym zadaniu:

```
# tracejob <numer_zadania>
```

### 1.22.2 Licencje Intel

Informacje dotyczące dostępnych i wypożyczonych licencji dostępne są na węźle `license`:

```
# service lmgrd_intel status
```

## 1.23 Zarządzanie repozytorium

### 1.23.1 Dodawanie paczek do repozytorium

W celu dodania paczek do repozytorium należy zalogować się na maszynę `nethost` i przejść do katalogu gdzie znajdują się paczki danego repozytorium np. PUIAS i pobrać interesującą nas paczkę

```
# cd /srv/repo/scientific-hpc-puias/Packages
# yumdownloader –disablerepo=* --enablerepo=PUIAS_6_computational
  <nazwa_paczki>
```

Następnie przejść do katalogu wyżej i zaktualizować repozytorium

```
# cd ../
# createrepo .
```

Paczki będą już dostępne na pozostałych węzłach. Należy pamiętać o wyczyszczeniu cachu repozytorium na węzłach

```
# yum clean all
```

### 1.23.2 Upgrade repozytorium

W celu wykonania aktualizacji repozytorium należy na węźle obliczeniowym przejść do katalogu `/root/scripts` i wykonać wywołać skrypt `sl-64-updates.lftp`

```
# cd /root/scripts
# lftp –f sl-64-updates.lftp
```

Paczki będą dostępne na węzłach po wyczyszczeniu pamięci cache.