Развёртывание высокопроизводительного кластера с использованием менеджера ресурсов Slurm и контейнеров Singularity

Выполнил студент группы
ИВТ-42-БО
Коляда М.М.

Научный руководитель
Доктор физ.-мат наук
Глызин С.Д.

ЯрГУ им. П. Г. Демидова

Постановка задачи

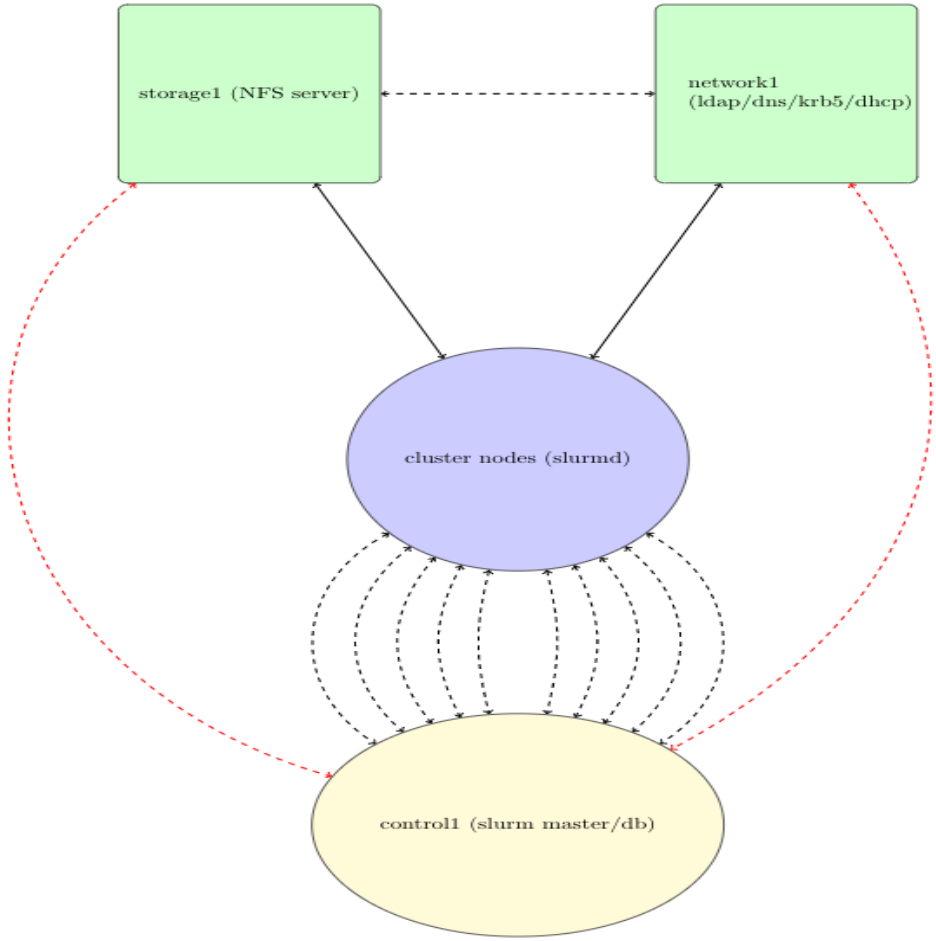
- Автоматизировать развёртывание кластера

Постановка задачи

- Автоматизировать развёртывание кластера
- Создать образы ОС для запуска вычислительных задач в контейнерах

Используемый стек

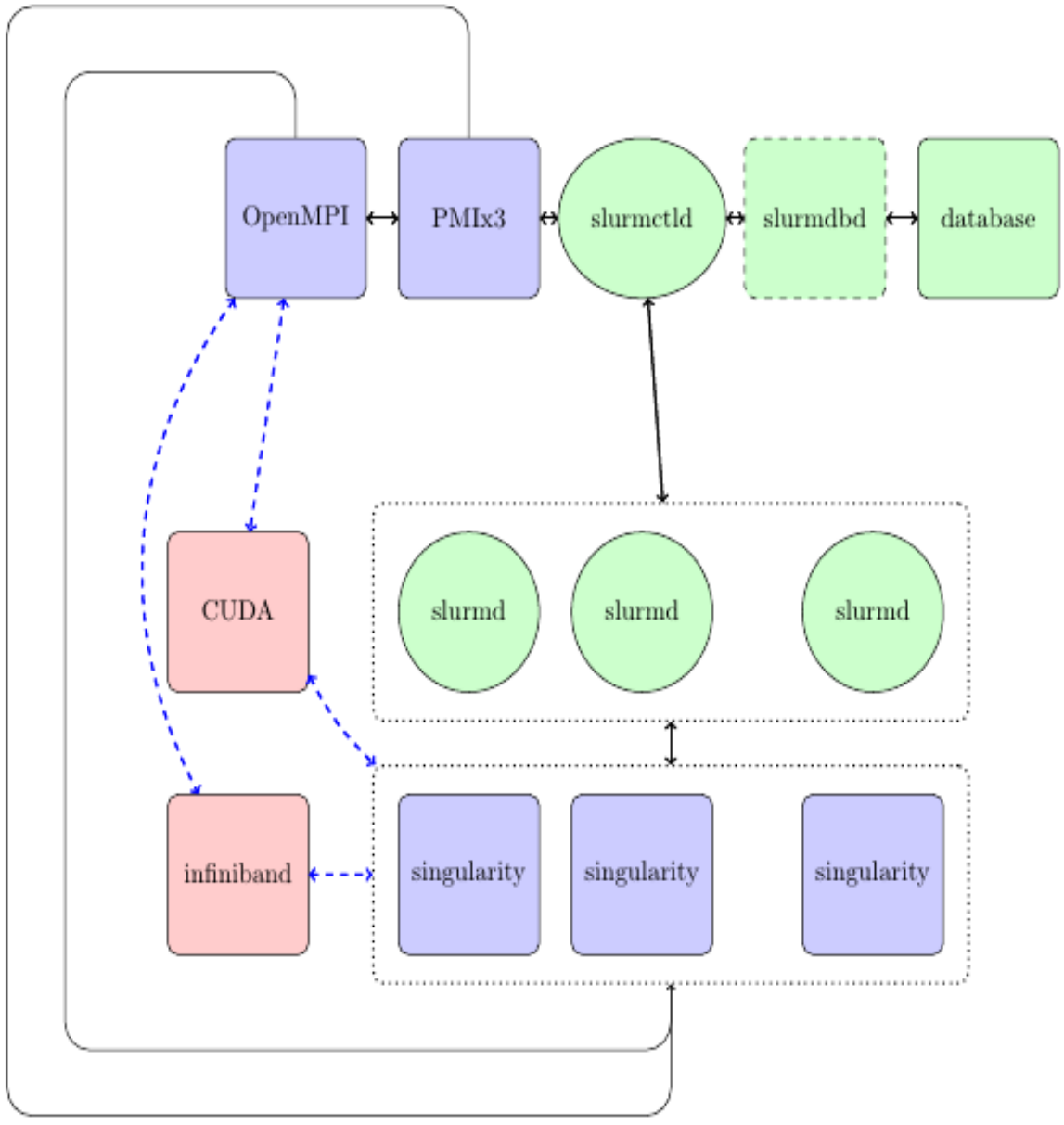
Архитектура и сервисы кластера


storage1 (NFS server)
network1
(ldap/dns/krb5/dhcp)
cluster nodes (slurmd)
control1 (slurm master/db)

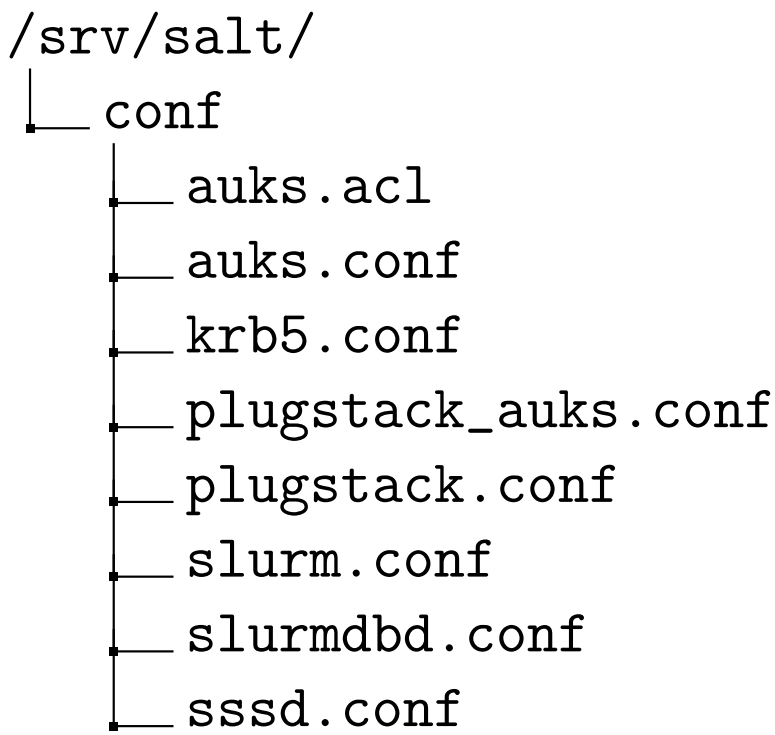
Пример пользователя (ldap)

```
dn: uid=ikrasnov,ou=people,dc=int,dc=corp7
changetype: add
objectClass: inetOrgPerson
objectClass: posixAccount
objectClass: shadowAccount
uid: ikrasnov
cn: Ignat Krasnov
displayName: Ignat Krasnov
sn: Krasnov
givenName: Ignat
uidNumber: 10025
gidNumber: 5000
loginShell: /bin/bash
homeDirectory: /clusterhome/ikrasnov
```

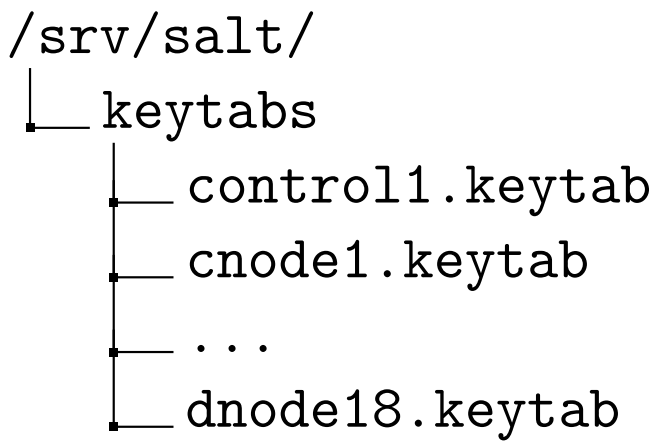
Архитектура кластера (ПО)


OpenMPI
PMIx3
slurmctld
slurmdbd
database
CUDA
slurmd
slurmd
slurmd
infiniband
singularity
singularity
singularity

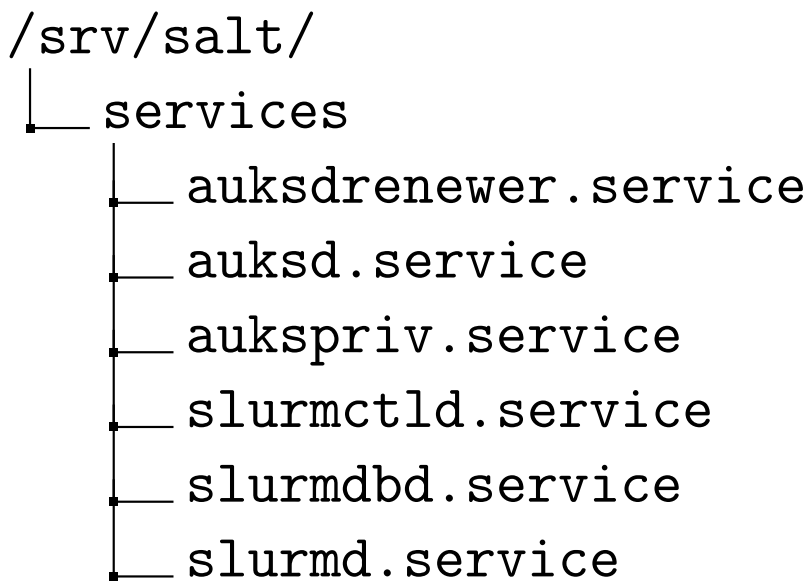
Структура корня saltstack



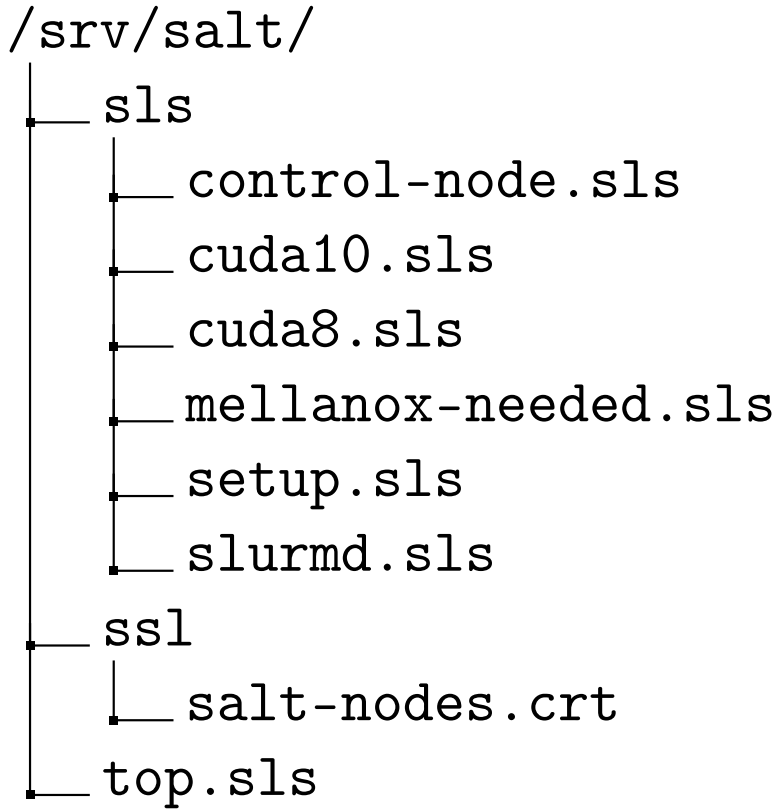
Структура корня saltstack (keytabs)



Структура корня saltstack (systemd services)



Структура корня saltstack (state файлы)



корневой файл saltstack (top.sls)

```
base:
    '*':
        - sls.setup
    'cnode(01|02|03|04|05|06|07).int.corp7':
        - match: pcre
        - sls.cuda8
    'control1.int.accelcomp.org':
        - match: pcre
        - sls.control-node
        - sls.cuda10
        - sls.mellanox-needed
       ....
    'dnode(09|10|11|12|13|14).int.corp7':
        - match: pcre
        - sls.mellanox-needed
    'not control1.int.corp7':
        - match: compound
        - sls.slurmd
```

Пример установки mellanox infiniband

```
mellanox specified packages:
    pkg.installed:
        - pkgs:
          - m4
          - swig
          - autoconf
          ...

unpack mlx_ofed:
    archive.extracted:
        - name: /usr/local
        - source:
        ↪  /clusterhome/install/saltcluster/MLNX_OFED_LINUX-4.6.tar.gz

install mlx_ofed:
    cmd.run:
        - name: perl /usr/local/MLNX_OFED/mlnxofedinstall --force
        - runas: root
        - unless:
            - ls /usr/sbin/iblinkinfo
```

Пример создания пользователя

```
slurm_user:
    user.present:
        - name: slurm
        - fullname: Slurm
        - shell: /bin/false
        - home: /home/slurm
        - uid: 64030
        - gid_from_name: True
        - require:
            - group: slurm
```

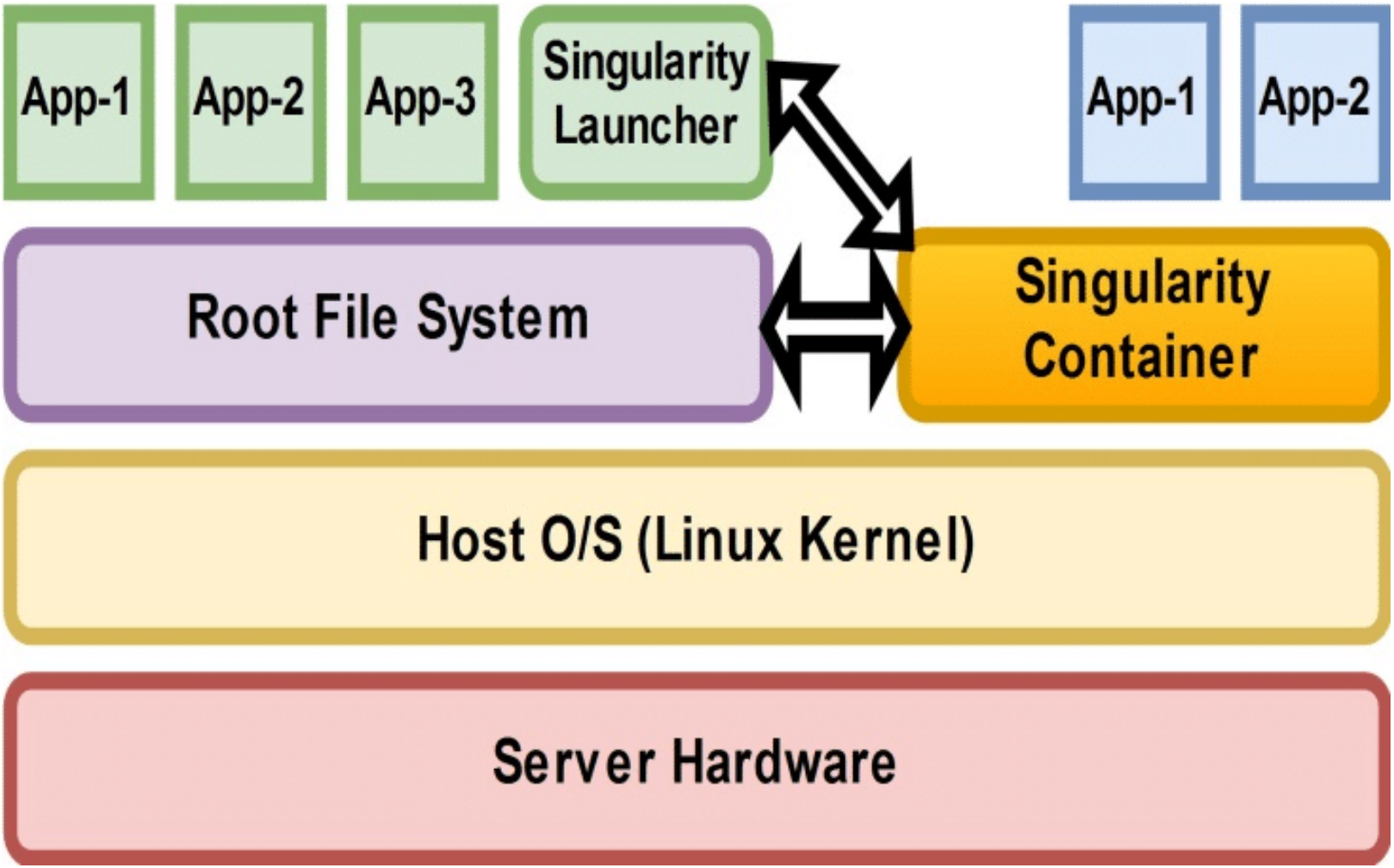
slurm.conf

```
SlurmctldHost=control1
MpiDefault=pmix_v3
PluginDir=/usr/local/slurm/lib/slurm
SlurmdPort=6818
SlurmdSpoolDir=/home/slurm/slurmd
SlurmUser=slurm
StateSaveLocation=/home/slurm
SlurmctldDebug=debug5
SlurmctldLogFile=/tmp/slurmctld.log
SlurmdDebug=debug5
SlurmdLogFile=/tmp/slurmd.log
...
NodeName=dnode[01-04] RealMemory=64000 Sockets=2 CoresPerSocket=10
↪  ThreadsPerCore=1
NodeName=dnode[05-08] RealMemory=128000 Sockets=2 CoresPerSocket=10
↪  ThreadsPerCore=1
NodeName=dnode[09-14] RealMemory=16000 Sockets=2 CoresPerSocket=6
↪  ThreadsPerCore=1
NodeName=cnode[1-7] RealMemory=64000 Sockets=2 CoresPerSocket=8
↪  ThreadsPerCore=1
PartitionName=all Nodes=cnode[1-7],dnode[01-14]
```

Процесс развёртывания

```
salt '*' state.apply && salt '*' system.reboot
```

Немного о singularity



Создание образа контейнера

```
FROM ubuntu:16.04

RUN apt-get update
RUN echo \"deb [trusted=yes] http://192.168.1.3 xenial main\" >> /etc/apt/source.list
# sync ppa in
RUN apt get-update
RUN DEBIAN_FRONTEND=noninteractive apt-get install -y libevent-*
RUN DEBIAN_FRONTEND=noninteractive apt-get install -y cuda-10.1
RUN DEBIAN_FRONTEND=noninteractive apt-get install -y openmpi4 pmix
```

slurm accounting

```
      User    Account  Partition                       QOS
---------- ---------- ---------- ------------------------------
   sglyzin    regular      debug    normal,short_term,standby
   zlogene      power               normal,standby
   zlogene    regular      debug    normal,short_term,standby
   zlogene    regular               normal,short_term,standby
   zlogene    regular        uni    normal,short_term,standby
```

Пример запуска задачи

$ srun -p uni -A power singularity exec
docker://storage1.int.accelcomp.org/dgl-ubuntu:latest hostname

```
Hello world from processor cnode01, rank 1 out of 4 processors
Hello world from processor cnode02, rank 0 out of 4 processors
Hello world from processor cnode03, rank 3 out of 4 processors
...
Hello world from processor cnode07, rank 2 out of 4 processors
```

Спасибо за внимание!