

Министерство образования и науки Российской Федерации
САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРВА ВЕЛИКОГО
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

ОТЧЕТ

по дисциплине «Методы управления киберфизическими системами на
суперкомпьютерах»

на тему «Установка и настройка SLURM»

Выполнила

студентка группы №3640201/90201

_____ Быкова К. А.

подпись

Проверил

кандидат технических наук

_____ Чуватов М. В.

подпись

Санкт-Петербург – 2021

Оглавление

Постановка задачи.....	3
SLURM.....	4
Архитектура SLURM.....	4
MUNGE	5
Установка и настройка SLURM	5
Конфигурация SLURM на управляющем узле.....	5
Конфигурация SLURM на исполнительных узлах	6
Проверка работоспособности SLURM	6
Запуск тестовой задачи в SLURM.....	6
Вывод	8
Список литературы	9

Постановка задачи

Установить и настроить менеджера кластеров и планировщика задач SLURM на нескольких узлах. Для проверки работоспособности выполнить тестовую задачу на узлах.

SLURM

SLURM (Simple Linux Utility for Resource Management)- это открытая, надежная и хорошо масштабируемая система управления ресурсами кластера с планировщиком задач, применяемая как для больших, так и для малых Linux-кластеров.

Как менеджер ресурсов кластера, SLURM выполняет три ключевых функции:

- Определяет выделенный и/или совместный доступ пользователей к ресурсам (вычислительным узлам) на некоторое время для выполнения ими вычислительных задач.
- Обеспечивает функционирование структуры запуска, выполнения и мониторинга задач (обычно это параллельные задачи) на выделенных узлах.
- Распределяет ресурсы, управляя очередью ожидающих запуска задач.

Архитектура SLURM

SLURM состоит из сервиса slurmd, запускающегося на каждом вычислительном узле, и центрального сервиса slurmctld, запускающегося на управляющем узле (опционально - с резервной копией управляющего узла).

Сервисы slurmd образуют отказоустойчивую иерархическую структуру.

Пользовательские команды включают: sacct, salloc, sattach, sbatch, sbcast, scancel, scontrol, sinfo, smap, squeue, srun, strigger и svview. Все эти команды могут быть запущены как с управляющего сервера, так и с узлов кластера (рисунок 1).

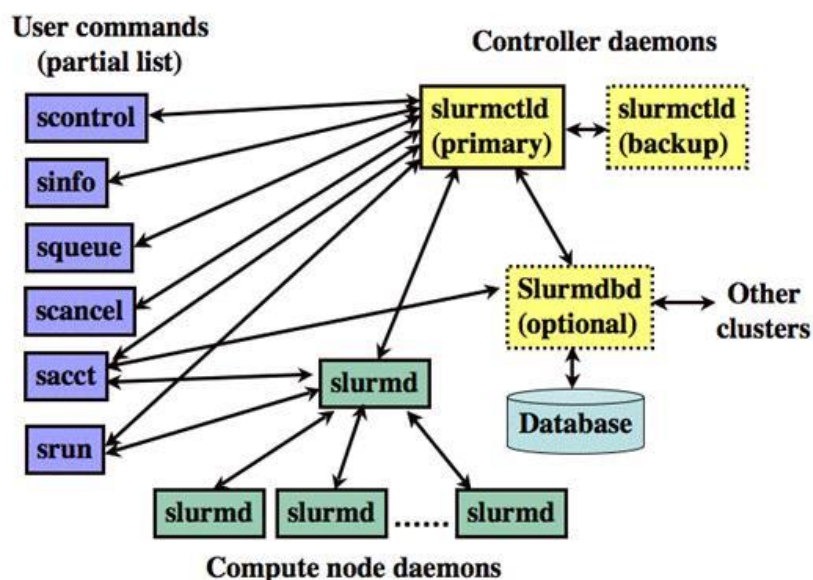


Рисунок 1. Схема архитектуры SLURM

Объекты, управляемые сервисами SLURM, это узлы - вычислительный ресурс SLURM-а, разделы, которые объединяют узлы в логические множества, задачи подмножества ресурсов, выделенных пользователю на указанное количество времени, и шаги задачи, которые являются множествами подзадач в рамках задачи.

Разделы могут рассматриваться как очереди задач, каждая из которых имеет комплекс ограничений, как например ограничение задачи по размеру, по времени выполнения, уровню доступа пользователей и т. д. Задачи в очереди упорядочены по приоритету и им выделяются ресурсы в соответствующем разделе.

Как только для задачи выделено множество узлов, пользователь может запускать параллельную работу в виде шагов задачи в любой конфигурации, в пределах выделенных узлов.

Для примера задача может быть запущена таким образом, что один единственный её шаг выполнения займет все выделенные ресурсы, в то же время несколько подзадач могут независимо друг от друга использовать часть выделенных ресурсов.

MUNGE

MUNGE (MUNGE Uid 'N' Gid Emporium) — это служба аутентификации для создания и проверки учетных данных. Он предназначен для высокой масштабируемости для использования в кластерной среде HPC. Он позволяет процессу аутентифицировать UID и GID другого локального или удаленного процесса в группе хостов, имеющих общих пользователей и группы. Эти узлы образуют область безопасности, которая определяется общим криптографическим ключом. Клиенты в этой области безопасности могут создавать и проверять учетные данные без использования привилегий суперпользователя, зарезервированных портов или специфичных для платформы методов.

Установка и настройка SLURM

Конфигурация SLURM на управляющем узле

Установка SLURM на управляющем узле выполняется командой:

```
apt install slurm-wlm
```

После установки необходимо произвести конфигурацию SLURM. Для этого был взят пример конфигурации из файла `/usr/share/doc/slurm-client/examples/slurm.conf.simple.gz` и помещен по пути `/etc/slurm-llnl/slurm.conf`.

В файле были изменены следующие параметры (представлен фрагмент кода):

```
ControlMachine=mgnt          #Имя управляющего узла
ControlAddr=192.168.122.10    #Адрес управляющего узла
NodeName=cn[01-02] Procs=1 State=UNKNOWN #Список вычислительных узлов
PartitionName=test Nodes=cn[01-02] Default=YES MaxTime=INFINITE State=UP
#Добавление вычислительных узлов в группу "test".
```

При установке SLURM устанавливается так же пакет MUNGE и происходит автоматическая генерация munge.key по пути /etc/munge/munge.key.

Конфигурация SLURM на исполнительных узлах

Для установки slurm на исполнительные узлы нужно выполнить следующую команду:

```
apt install slurmd
```

После этого необходимо скопировать файлы /etc/slurm-llnl/slurm.conf и /etc/munge/munge.key на исполнительные узлы.

Эти 2 шага выполняются в автоматическом режиме на 2-х узлах с помощью Ansible.

Проверка работоспособности SLURM

После настройки главного и исполнительных узлов проверить доступность можно с главного узла с помощью команды: *sinfo* (представлен фрагмент кода):

```
root@mgnt:~# sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
debug*      up   infinite    2  down cn[01-02]
```

Если узлы указаны как down или unk, то можно задать их доступность командой:

```
scontrol update nodename=cn01 state=idle
```

Запуск тестовой задачи в SLURM

Для запуска задач в SLURM существует 2 команды: *sbatch* и *srun*. Их отличие заключается в том, что *srun* запускает задачу сразу же и ожидает результат, а *sbatch* добавляет команду в очередь.

Уже запущенные и ждущие выполнения задачи можно смотреть через команду *squeue*, а результат работы уже выполненных задач командой *scontrol show job*.

Для проверки работы SLURM был написан короткий скрипт:

```
#!/bin/bash
sleep 2
echo "test"
```

Скрипт ожидает 2 секунды чтобы симитировать произведение вычислений и выводит строку как результат.

Скрипт запущен для демонстрации работы очереди. Вывод команды *squeue* представлен фрагмент кода:

```
root@mgnt:~# squeue
      JOBID PARTITION  NAME  USER ST  TIME  NODES
NODELIST(REASON)
      3    debug script.s  root PD   0:00    1 (Resources)
```

После выполнения задач можно выполнить команду: *scontrol show job*.

Результат работы скрипта сохранен на узле cn01, представлен фрагмент кода:

```
root@cn01t:~# ls
post.log  post.script  slurm-2.out  slurm-1.out
root@cn01t:~# tail slurm-2.out
test
```

Вывод

В ходе работы был установлен и сконфигурирован SLURM для запуска тестовых задач на узлах cn01 и cn02. Для проверки написан короткий скрипт. Проверка работы скрипта выполнена через SLURM и просмотр файла с результатом работы на узле cn01.

Список литературы

1. Slurm - Wikipedia [Электронный ресурс] : Wikipedia. The Free Encyclopedia. -М ., 2001-2021. - Режим доступа: https://en.wikipedia.org/wiki/Slurm_Workload_Manager (дата обращения 17.04.2021);
2. Munge - GitHid [Электронный ресурс] : GitHub. -М., 2021. - Режим доступа: <https://github.com/dun/munge/wiki> (дата обращения 17.04.2021);
3. Администрирование кластера [Электронный ресурс] : Режим доступа: <https://parallel.uran.ru/book/export/html/513> (Дата обращения: 17.04.2021).