



МИНОБРНАУКИ РОССИИ

ФИЛИАЛ

**ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО УЧРЕЖДЕНИЯ «ФЕДЕРАЛЬНЫЙ НАУЧНЫЙ ЦЕНТР НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ
ИНСТИТУТ СИСТЕМНЫХ ИССЛЕДОВАНИЙ РОССИЙСКОЙ АКАДЕМИИ НАУК»**

**Межведомственный суперкомпьютерный центр Российской академии наук
(МСЦ РАН – филиал ФГУ ФНЦ НИИСИ РАН)**

УТВЕРЖДАЮ

**Директор МСЦ РАН – руководитель
ЦКП вычислительными ресурсами
МСЦ РАН – филиала ФГУ ФНЦ
НИИСИ РАН**



Б.М. Шабанов

«26» 12 2019 г.

Технический отчет

**Интеграция системы доступа к оборудованию Центра с
цифровой системой управления сервисами научной
инфраструктуры коллективного пользования (УСНИКП)**

Технический отчет рассмотрен и одобрен на заседании Ученого
совета МСЦ РАН – секции Ученого совета ФГУ ФНЦ НИИСИ РАН

«25» 12 2019 г., протокол № 6

Москва 2019

Содержание

1 Интеграция системы доступа к оборудованию Центра с цифровой системой управления сервисами научной инфраструктуры коллективного пользования (УСНИКП).....	3
1.1 Перечень документов, на основании которых осуществлялась интеграция системы доступа в УСНИКП.....	3
1.2 Цель и состав выполненных задач по интеграции системы доступа к оборудованию Центра с УСНИКП.....	4
1.3 Выделение и предоставление разделов суперкомпьютеров ЦКП для размещения заданий системы УСНИКП.....	5
1.4 Сетевая интеграция выделенных разделов с системой УСНИКП....	6
1.5 Конфигурация и настройка применяемой в ЦКП системы коллективного пользования и управления заданиями для обеспечения возможности приема заданий из системы УСНИКП и их обработки на выделенных разделах.....	7
1.6 Программная документация для интегрированной системы.....	10
Перечень терминов.....	12
Перечень сокращений.....	13

1. Перечень документов, на основании которых осуществлялась интеграция системы доступа в УСНИКП

1.1 Приказ Министерства науки и высшего образования Российской Федерации № 394 от 19 июня 2019 года «О создании цифровой системы управления сервисами научной инфраструктуры коллективного пользования (в том числе ЦКП, УНУ), предоставляющей безбарьерный доступ исследователям с заказом услуг с использованием инфраструктуры, в том числе к оцифрованным коллекциям и банкам данных организаций, выполняющих научные исследования и разработки, а также отказ от излишней

бюрократизации, упрощение закупок материалов и образцов для исследований и разработок».

1.2 «Концепция цифровой автоматизированной системы управления сервисами научной инфраструктуры коллективного пользования», согласованная Протоколом №1 заседания Совета Министерства науки и высшего образования Российской Федерации по цифровому развитию и информационным технологиям (секция по цифровым платформам и единому цифровому пространству в сфере науки и высшего образования Российской Федерации) от 11 марта 2019 года.

1.3 Техническое задание на разработку цифровой системы управления сервисами научной инфраструктуры коллективного пользования (в том числе ЦКП, УНУ), предоставляющей безбарьерный доступ исследователям с заказом услуг с использованием инфраструктуры, в том числе к оцифрованным коллекциям и банкам данных организаций, выполняющих научные исследования и разработки, а также отказ от излишней бюрократизации, упрощение закупок материалов и образцов для исследований и разработок (АС УСНИКП), утвержденное Заместителем Министра науки и высшего образования Российской Федерации 08 июля 2019 г. (далее – ТЗ на АС УСНИКП).

1.4 Техническое задание на проведение работ по теме «Поддержка и развитие Центра коллективного пользования вычислительными ресурсами МСЦ РАН - филиала ФГУ ФНЦ НИИСИ РАН для обеспечения реализации приоритетов научно-технологического развития Российской Федерации», определенное Соглашением с Министерством науки и высшего образования Российской Федерации о предоставлении из федерального бюджета грантов в форме субсидий в соответствии с п. 4 статьи 78.1 Бюджетного кодекса Российской Федерации от 05.11.2019 № 075-15-2019-1636

2 Цель и состав выполненных задач по интеграции системы доступа к оборудованию Центра с УСНИКП

2.1 Целью интеграции системы доступа к оборудованию Центра с УСНИКП является предоставление цифровым платформам управления научными сервисами: УСНИКП, СУС (Система управления сервисами, разработанная ФИЦ ИУ РАН) - интерфейса для выделения и учета ресурсов, предоставляемых центром. Интерфейс предоставляется в локальной сети организации, ведущей разработку (УСНИКП, ИСП РАН) или опытную эксплуатацию (СУС, ФИЦ ИУ РАН) платформы.

2.2 Для осуществления интеграции были выполнены следующие задачи.

2.2.1 Выделение и предоставление разделов суперкомпьютеров ЦКП для размещения заданий автоматизированной системы УСНИКП.

2.2.2 Сетевая интеграция выделенных разделов с автоматизированной системой УСНИКП.

2.2.3 Конфигурация и настройка применяемой в ЦКП системы коллективного пользования и управления заданиями для обеспечения возможности приема заданий из автоматизированной системы УСНИКП и их обработки на выделенных разделах.

2.3 Разработка программной документации для интегрированной системы.

2.4 Выделение и предоставление разделов суперкомпьютеров ЦКП для размещения заданий системы УСНИКП

2.5 Для размещения заданий автоматизированной системы УСНИКП были выделены следующие разделы суперкомпьютеров ЦКП вычислительными ресурсами МСЦ РАН (ЦКП ВР МСЦ РАН).

- раздел суперкомпьютера МВС-100К;

- разделы KNL и Haswell суперкомпьютера MBC-10П ОП.

2.6 Раздел суперкомпьютера MBC-100К производительностью 12,1 TFlops включает 126 узлов на базе 4-ядерных процессоров Intel Xeon E5450 (Harpertown) 3,0 ГГц, объединенных коммуникационной средой Infiniband DDR. Каждый узел содержит 2 процессора Intel Xeon E5450 и 8 Гб оперативной памяти.

2.7 Раздел суперкомпьютера MBC-10П ОП Haswell производительностью 48,92 TFlops включает 42 узла на базе 14-ядерных процессоров Intel Xeon Haswell, объединенных коммуникационной средой Intel Omni-Path. Каждый узел включает в себя 2 процессора Intel Xeon E5-2697v3 и 128 Гб оперативной памяти.

2.8 Раздел KNL производительностью 131,33 TFlops включает 38 ВМ на базе 72-ядерных процессоров Intel Xeon Phi 7290 (KNL), объединенных коммуникационной средой Intel Omni-Path. Каждый узел включает в себя процессор Intel Xeon Phi 7290 (KNL) 1,5 ГГц и 96 Гб оперативной памяти.

2.9 В качестве программной управляющей системы на всех разделах установлена Система управления прохождением параллельных заданий (СУППЗ).

3 Сетевая интеграция выделенных разделов с системой УСНИКП

3.1 Сетевая интеграция с автоматизированной системой (АС) УСНИКП осуществляется на базе Национальной исследовательской компьютерной сети.

3.2 Для работы АС УСНИКП и выделенными ресурсами ЦКП ВР МСЦ РАН организуются виртуальные сети:

- доступа;
- транспортная

3.3 Сеть доступа – сеть L2 /24 с приватным адресным пространством между сервером АС УСНИКП и серверами доступа выделенных для запуска

задач АС УСНИКП разделов суперкомпьютеров ЦКП ВР МСЦ РАН. Для серверов доступа разделов суперкомпьютерных ресурсов ЦКП ВР МСЦ РАН сеть доступа является доверенной, для компьютеров в данной сети разрешен вход по протоколу ssh для запуска заданий. Маршрутизация других сетей в сеть доступа отсутствует.

3.4 Транспортная сеть – сеть L2 /24 с приватным адресным пространством между сервером системы хранения данных АС УСНИКП и локальным репозиторием контейнеров (Docker репозиторий, см. раздел 1.5). Для сервера – локального репозитория контейнеров транспортная сеть является доверенной, для компьютеров в данной сети разрешен вход по протоколу ssh для копирования контейнеров. Маршрутизация других сетей в сеть доступа отсутствует.

3.5 Для сервера локального репозитория контейнера допустимо монтирование сетевой файловой системы nfs с серверов систем хранения данных АС УСНИКП при условии, что права доступа к файлам определяются в соответствии с учетными записями пользователей ЦКП ВР МСЦ РАН. Синхронизация или трансляция учетных записей ЦКП ВР МСЦ РАН в учетные записи АС УСНИКП осуществляется средствами АС УСНИКП.

3.6 Виртуальные сети (сеть доступа и транспортная) проложены через четыре домена: локальную сеть МСЦ РАН, Опорную телекоммуникационную сеть (ОТС) РАН в составе Национальной исследовательской компьютерной сети, локальную сеть ИСП РАН, локальную сеть ФИЦ ИУ РАН. Соответствующие vlan id 811 (сеть доступа), 812 (транспортная сеть) согласованы сетевыми (системными) администраторами локальных сетей МСЦ РАН, ИСП РАН, ФИЦ ИУ РАН, Центром управления сетью ОТС РАН.

4 Конфигурация и настройка применяемой в ЦКП системы коллективного пользования и управления заданиями для обеспечения возможности приема заданий из системы УСНИКП и их обработки на выделенных разделах

4.1 В качестве системы коллективного пользования и управления заданиями на всех выделенных для размещения заданий УСНИКП разделах суперкомпьютеров ЦКП ВР МСЦ РАН установлена Система управления прохождением параллельных заданий (СУППЗ).

4.2 Конфигурация и настройка СУППЗ заключалась в изменении в соответствии с Руководством (см. п. 5.3) управляющих командных файлов – сценариев запуска и завершения задания для обеспечения возможности в соответствии с п. 4.12.2.1 ТЗ на АС УСНИКП сервиса предоставления виртуальных машин по запросу.

4.3 Для возможности запуска заданий из АС УСНИКП, представленных в виде контейнеров, в ЦКП ВР МСЦ РАН был организован локальный Docker-репозиторий, схема которого представлена на рисунке 1.

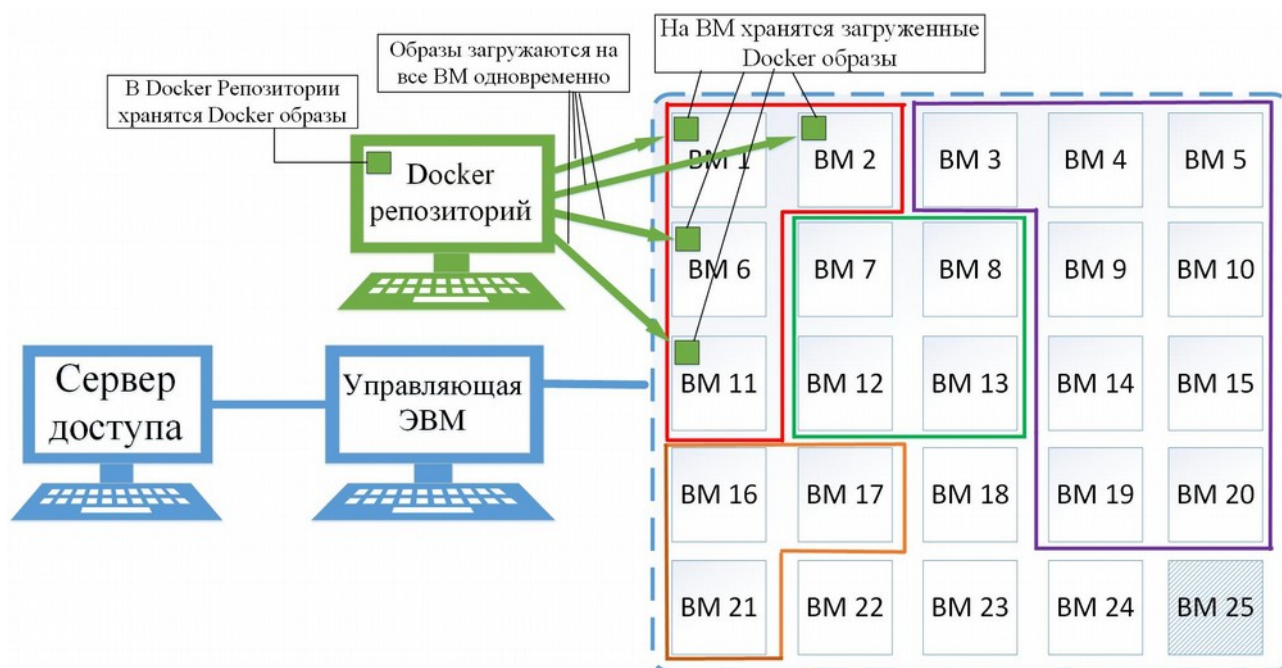


Рисунок 1 – Схема Docker-репозитория ЦКП ВР МСЦ РАН

Для работы СУППЗ с заданиями, представленными в виде docker-контейнеров, необходимо организовать следующие операции:

- загрузка docker-образов на узлы (вычислительные модули – VM) суперкомпьютера и их конфигурирование;
- запуск задания в контейнерах на выделенных VM;
- остановка контейнеров после завершения задания и их дальнейшее удаление с VM.

4.4 Загрузка docker-образов на VM и их конфигурирование может быть проведено на стадии подготовки VM, поэтому было принято решение реализовать этот функционал в сценарии СУППЗ `makeup_node`. Остановка и дальнейшее удаление контейнеров с VM должны быть проведены по завершению задания, этого можно добиться, внося изменения в сценарий СУППЗ `cleanup_node`. Сценарии `makeup_node` и `cleanup_node` рекомендованы к изменению системным программистом, поэтому внесение в них изменений является типичной операцией для СУППЗ.

4.5 Для развертывания docker-репозитория был выделен отдельный узел, к которому имеют доступ все VM высокопроизводительного кластера. Запуск docker-репозитория производится путем запуска специального контейнера из образа под названием “registry:v2”. Образ “registry:v2” может быть загружен на узел с помощью команды

```
docker pull registry:v2
```

или при запуске контейнера с указанием данного образа:

```
docker run registry:v2
```

4.6 Для исключения возможности подмены пользователями загруженных в репозиторий образов на собственные и в дальнейшем их использования на VM, существует возможность авторизации в репозитории с помощью сертификатов, которые могут быть созданы с использованием `openssl`, или с помощью логина и пароля.

4.7 Для загрузки новых образов в репозиторий администратору необходимо использовать команду

```
docker push reposAddr:port/ImageName
```

где `reposAddr` – адрес репозитория (ip-адрес, доменное имя), `port` – порт который открыт для работы с репозитория на узле с адресом `reposAddr`, `imageName` – имя образа. Перед загрузкой администратцию необходимо пройти авторизацию в репозитории с помощью команды

```
docker login
```

4.8 Развертывание задания в контейнерах на выделенных ВМ логично вынести в сценарий `runmvs.bat`, который запускается от имени пользователя на первом из выделенных ВМ. Однако, работа с контейнерами подразумевает работу от имени суперпользователя, поэтому было решено внести изменения в сценарий СУППЗ `runtask`. В видоизмененном командном файле `runtask` от имени суперпользователя производится запуск сценария по развертыванию задания в контейнерах.

4.9 Все ВМ раздела суперкомпьютера находятся в одной сети с репозиториум и могут загрузить из него необходимый docker-образ. Для загрузки образа на ВМ на нем должна быть выполнена команда:

```
docker pull repository:port/ImageName
```

4.10 После выполнения команды начинается загрузка образа. Загрузка может производиться несколькими узлами одновременно, данную процедуру целесообразно проводить на стадии подготовки узлов к выполнению задания (сценарий `makeup_node` в СУППЗ).

4.11 Для взаимодействия с docker-репозиториум в СУППЗ должна храниться информация об адресе репозитория, именах, загружаемых из него образов, и сертификатов (или логина и пароля) для авторизации в репозитории. В СУППЗ присутствует возможность хранения указанной информации в

паспорте задания СУППЗ. В формируемый АС УСНИКП паспорт задания была добавлена секция `Docker`, в которой АС УСНИКП может указать тип контейнера (тип используется только внутри сценариев СУППЗ), название образа и другую сопутствующую информацию. Имя репозитория определяется в конфигурационном файле СУППЗ (`/usr/runmvs/.grunmvs`). В этом конфигурационном файле также хранится местонахождение сценариев, рассмотренных в пп. 4.4, 4.8.

5 Программная документация для интегрированной системы

5.1 Для интегрированной системы разработана программная документация в следующем составе:

- Методические рекомендации по использованию оборудования и программного обеспечения Центра коллективного пользования вычислительными ресурсами МСЦ РАН – филиала ФГУ ФНЦ НИИСИ РАН (далее – методические рекомендации);
- Руководство по конфигурации и настройке СУППЗ.

5.2 Разработанные методические рекомендации включают в себя командный интерфейс (CLI) программиста, позволяющий направить задание из АС УСНИКП на выполнение на любой из выделенных разделов суперкомпьютеров Центра.

5.3 Руководство по конфигурации и настройке СУППЗ позволяет изменять поведение системы в момент запуска задания из облачной среды АС УСНИКП и запускать задание в виде контейнеров `Docker` или виртуальных машин.

Перечень терминов

В настоящем документе применены следующие термины с соответствующими определениями:

параллельная программа: программа, состоящая из нескольких взаимодействующих процессов, которые могут одновременно выполняться на разных процессорах;

задание (параллельное задание): информационный объект, в состав которого входят:

- параллельная программа, решающая прикладную задачу;
- паспорт задания;
- входные данные параллельной программы.

паспорт задания: информационный объект, представленный в виде файла и содержащий описание параллельного задания и требования к вычислительным ресурсам.

Перечень сокращений

АС УСНИКП	– автоматизированная система управления сервисами научной инфраструктуры коллективного пользования
ВМ	– вычислительный модуль
ОС	– операционная система
СУППЗ	– система управления прохождением параллельных заданий
СУС	- система управления сервисами
УНУ	– уникальная научная установка
ЦКП	– центр коллективного пользования
ЦКП ВР МСЦ РАН	– Центр коллективного пользования вычислительными ресурсами МСЦ РАН – филиала ФГУ ФНЦ НИИСИ РАН

ПОДГОТОВЛЕНО:

Ведущий научный сотрудник А.В. Баранов

Научный сотрудник Д.В. Вершинин



СОГЛАСОВАНО:

Заместитель директора А.П. Овсянников

Заместитель директора П.Н. Телегин

Главный инженер В.М. Опалев

