Облако ИСП РАН

наш маленький Amazon EC2

Борисенко Олег

ИСП РАН

Цели проекта

- Развернуть облачную среду для использования сотрудниками института
- Минимизировать расходы на Амазон
- Предоставить возможность быстро и легко считать задачи над большими данными
- Иметь возможность сравнительно быстро модифицировать и обновлять систему
- Разобраться с тем, как вообще работают и устроены облачные системы на живом примере

Этапы проекта

- Проектирование и подбор железа
- Закупка и ожидание железа
- Реформирование серверной под энергетические требования
- Коммутация
- Разработка автоматических сценариев для разворачивания Openstack из исходников
- Настройка, борьба с ошибками (как Openstack, так и нашими)
- Анализ реальной производительности систем хранения

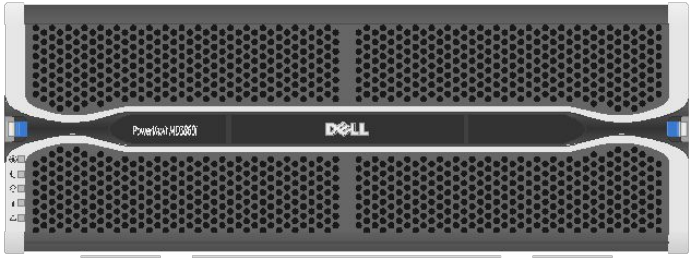
Суммарная мощность системы

Характеристики:

- 3968GB RAM без учета overcommit
- 256 ядер (512 с учетом HT)
- 28.5TB локальных для виртуальных машин дисков
- >130TB нелокального хранения с пропускной способностью 10Гбит/с
- Полная связность 20Гбит/с
- Канал на вход/выход 1Гбит/с, но будет больше
- 240 внешних адресов
- 1000 внутренних адресов в сети ИСП
- Энергопотребление под полной нагрузкой влезает в 65% доступной мощности

Железо/хранение

Dell PowerVault MD3820i
2U, 600W, 2047 BTU/час
24*1.2TB, SAS, 10k RPM ~22TB чистого места



Dell PowerVault MD3860i
4U, 1755W, 5988 BTU/час
60*4TB, SAS, 7.2k RPM ~200TB чистого места

Железо/вычисления

10U, 11000 W максимум по блокам питания,
7200 W максимум фактический

8x Dell M620:
2x Intel Xeon E5-2650 v2 (8 cores, 2.6Ghz)
256GB RAM (1866MHz ECC)
2x 10Gbit/s network

2x Dell MXL 10/40 Gbit/s:
2x 40Gbit/s
4x 10Gbit/s SFP+
4x 10Gbit/s Cat6 (витая пара)
Встроенное подключение всех серверов в корзине

10U, 11000 W максимум по блокам питания,
7200 W максимум фактический

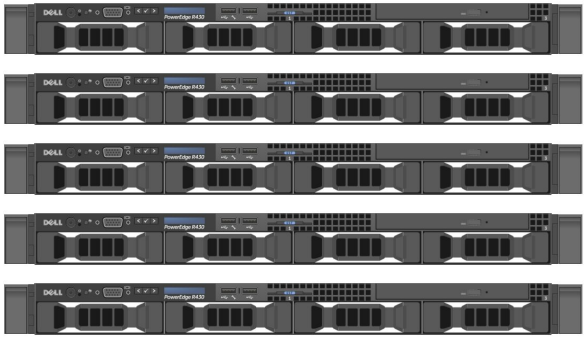
8x Dell M620:
2x Intel Xeon E5-2650 v2 (8 cores, 2.6Ghz)
256GB RAM (1866MHz ECC)
4x 10Gbit/s network

2x Dell MXL 10/40 Gbit/s:
Встроенное подключение всех серверов в корзине
2x 40Gbit/s
4x 10Gbit/s SFP+
4x 10Gbit/s Base-T (витая пара Cat6)

Железо/контроллеры

- 7x Supermicro, самосборные
- 1U, 560W
- 4x 1Tb SATA 7200RPM

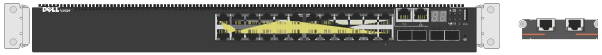
128GB RAM ECC, 1333Mhz
2x Intel Xeon E5-2670, 2.6Ghz, 8 ядер
2x 10Gbit/s Base-T (витая пара Cat6)
3x 1Gbit/s Base-T



32GB RAM, 1333Mhz
2x Intel Xeon E5-2620, 2.0Ghz, 6 ядер
2x 10Gbit/s Base-T
3x 1Gbit/s Base-T

32GB RAM, 1333Mhz
2x Intel Xeon E5-2620, 2.0Ghz, 6 ядер
2x 10Gbit/s Base-T
2x 10Gbit/s SFP+
1x 1Gbit/s Base-T

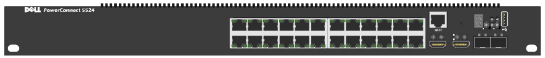
Железо/сеть



Dell N3024, 1U, 200W
24x 1Gbit/s
2x 1Gbit SFP
2x 10Gbit/s SFP+
2x 10Gbit/s Base-T (витая пара Cat6)

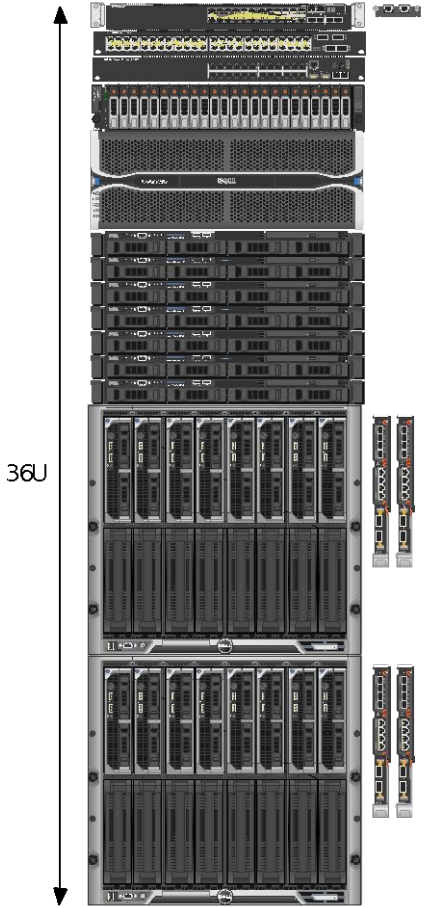
Dell S4820T, 1U, 460W
48x 10Gbit/s Base-T (витая пара Cat6)
4x 40Gbit/s SFP+

4x Dell MXL 10/40 Gbit/s:
Встроенное подключение всех серверов в корзине
2x 40Gbit/s
4x 10Gbit/s SFP+
4x 10Gbit/s Base-T (витая пара Cat6)



Dell PowerConnect 5524, 1U, 30W
24x 1Gbit/s Base-T (витая пара Cat5e)
2x 10Gbit/s SFP+

Энергопотребление системы



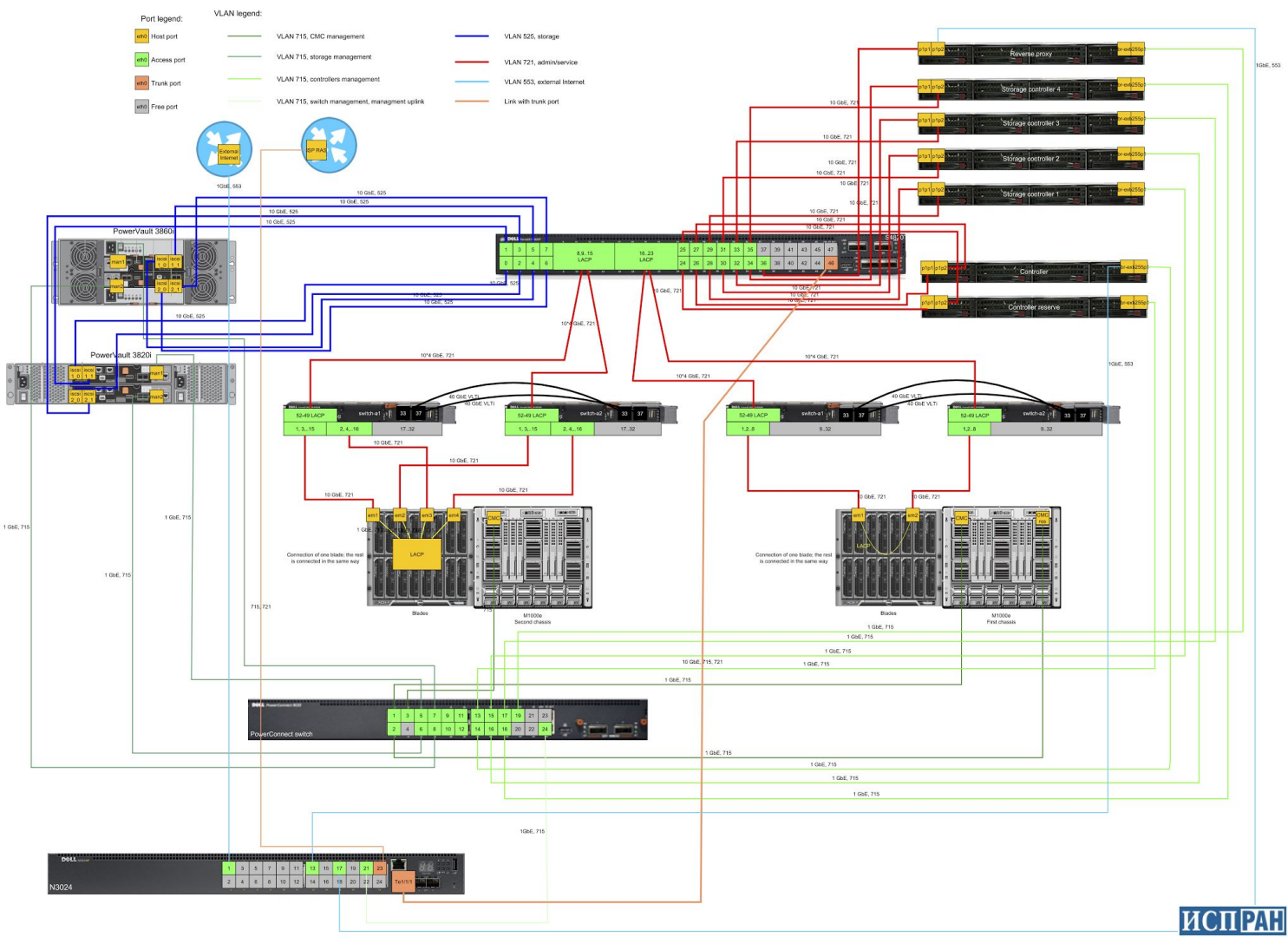
200W
460W
30W
600W
1755W
560W
560W
560W
560W
560W
560W
560W
7200W (11000W)
7200W (11000W)

21365W total max
(**28965**W)

Сетевая часть

Векторная версия с увеличением тут:

https://drive.google.com/file/d/0B4MHBKKyFpJfT3V3bmpyb29MUlE/view?usp=sharing



Хочу использовать, но не понимаю, зачем!

Краткий список того, что можно хотеть от облака:

- посчитать задачу, для которой мне не хватает памяти (нужно 128ГБ, например)
- считать одну и ту же задачу много раз и как-нибудь автоматически
- поднять систему сборки/CIS, и не хочу тратить ресурсы на поддержание воркеров
- протестировать принципиально новую ШикOS, но лень настраивать себе KVM
- посчитать задачу на Hadoop/Spark/etc на 100 узлов по 16 гигабайт памяти
- перестать думать о том, что куча мелких файлов не влезает на один жесткий диск, хочу кидать их куда-нибудь по одному, чтобы потом к ним обращаться по сети через REST API
- как-нибудь несложно поднять веб-сервис, который будет превращать любую картинку в котиков
- попробовать использовать скрипты на boto для Амазона, но бесплатно*
- эластично увеличивать число узлов в зависимости от загрузки уже существующих**

* совместимость далеко не полная, так что только попробовать

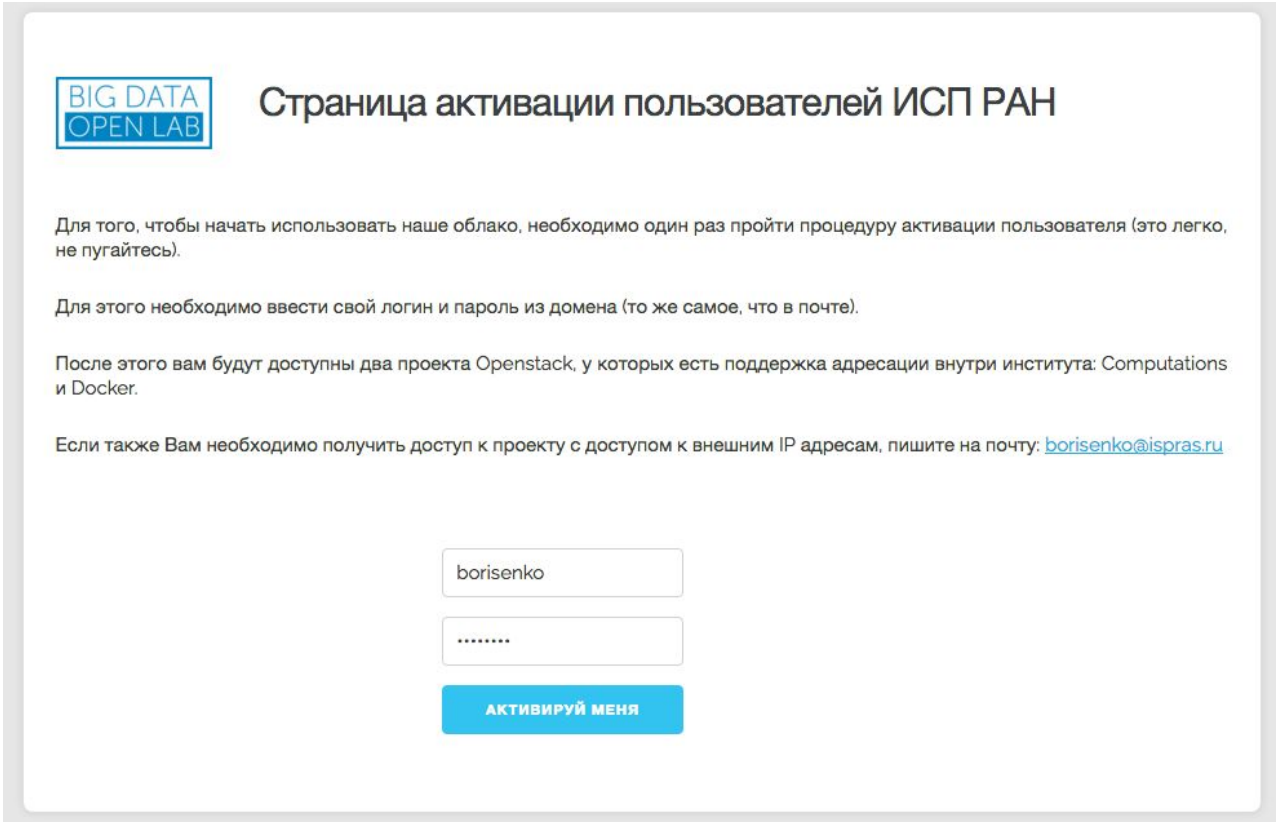
**требуются специальные навыки, это происходит не само собой

Как пользоваться

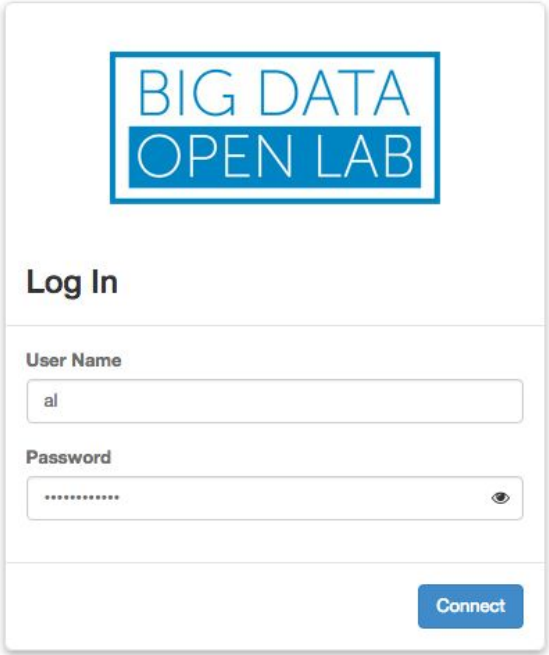
Прежде всего, необходимо себя активировать

1. Для этого нужно, чтобы были активным пользователем домена ИСП РАН (если у вас работает почта @ispras.ru, то это вы)
2. Нужно зайти на https://cloud.ispras.ru/ispusers/activation
   Эта ссылка откроется только через сети ИСП РАН
3. Вводите свой логин и пароль от почты и нажимаете “активировать”
4. Вас перебросит на страницу логина, тут нужно ввести логин и пароль еще раз

Активация


BIG DATA
OPEN LAB
Страница активации пользователей ИСП РАН
Для того, чтобы начать использовать наше облако, необходимо один раз пройти процедуру активации пользователя (это легко, не пугайтесь).
Для этого необходимо ввести свой логин и пароль из домена (то же самое, что в почте).
После этого вам будут доступны два проекта Openstack, у которых есть поддержка адресации внутри института: Computations и Docker.
Если также Вам необходимо получить доступ к проекту с доступом к внешним IP адресам, пишите на почту: borisenko@ispras.ru
borisenko
АКТИВИРУЙ МЕНЯ

Вход



Доступные образы

Вы можете добавлять и свои — это те, которые мы подготовили.

Требование: в образе должен быть зашит cloud-init.

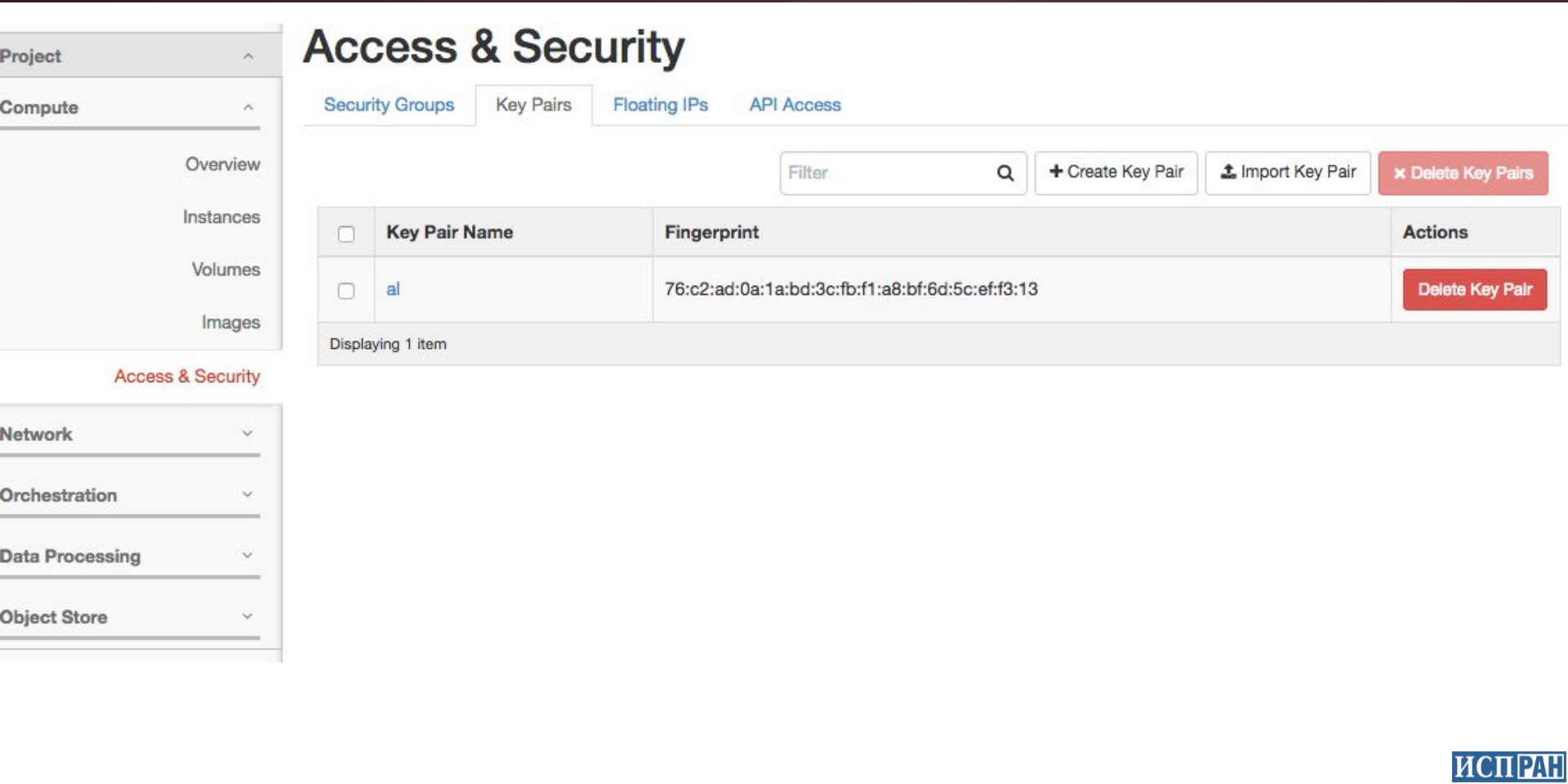
Images

Project (8) | Shared with Me (0) | Public (1) | + Create Image | × Delete Images

Image Name	Type	Status	Public	Protected	Format	Size	Actions
Ubuntu 12.04.5 for Cloudera 5.4.0 Sah...	Image	Active	No	No	QCOW2	3.1 GB	Launch Instance
CentOS 6.7 for Cloudera 5.4.0 Sahara ...	Image	Active	No	No	QCOW2	3.0 GB	Launch Instance
Ubuntu 14.04.4 Trusty x86_64 cloud image	Image	Active	No	No	Raw	247.4 MB	Launch Instance
Ubuntu 12.04.5 Precise x86_64 cloud i...	Image	Active	No	No	Raw	251.4 MB	Launch Instance
Debian 8.4 Jessie x86_64 cloud image	Image	Active	No	No	QCOW2	458.8 MB	Launch Instance
Fedora 23 x86_64 cloud image	Image	Active	No	No	QCOW2	223.5 MB	Launch Instance
CentOS 7 x86_64 cloud image	Image	Active	No	No	QCOW2	872.9 MB	Launch Instance
CentOS 6.7 x86_64 cloud image	Image	Active	No	Yes	QCOW2	712.3 MB	Launch Instance

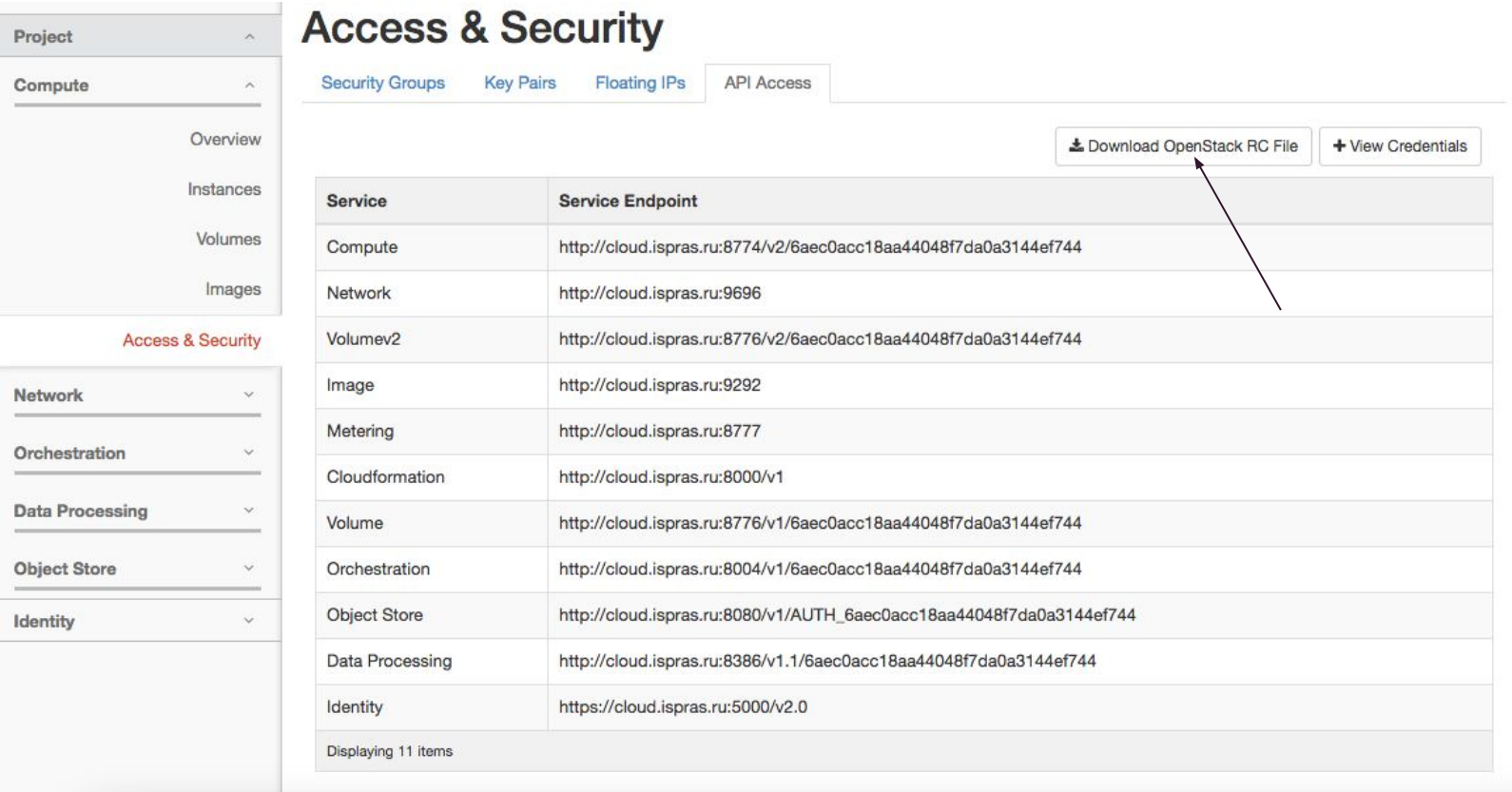
Displaying 8 items

Запуск машины

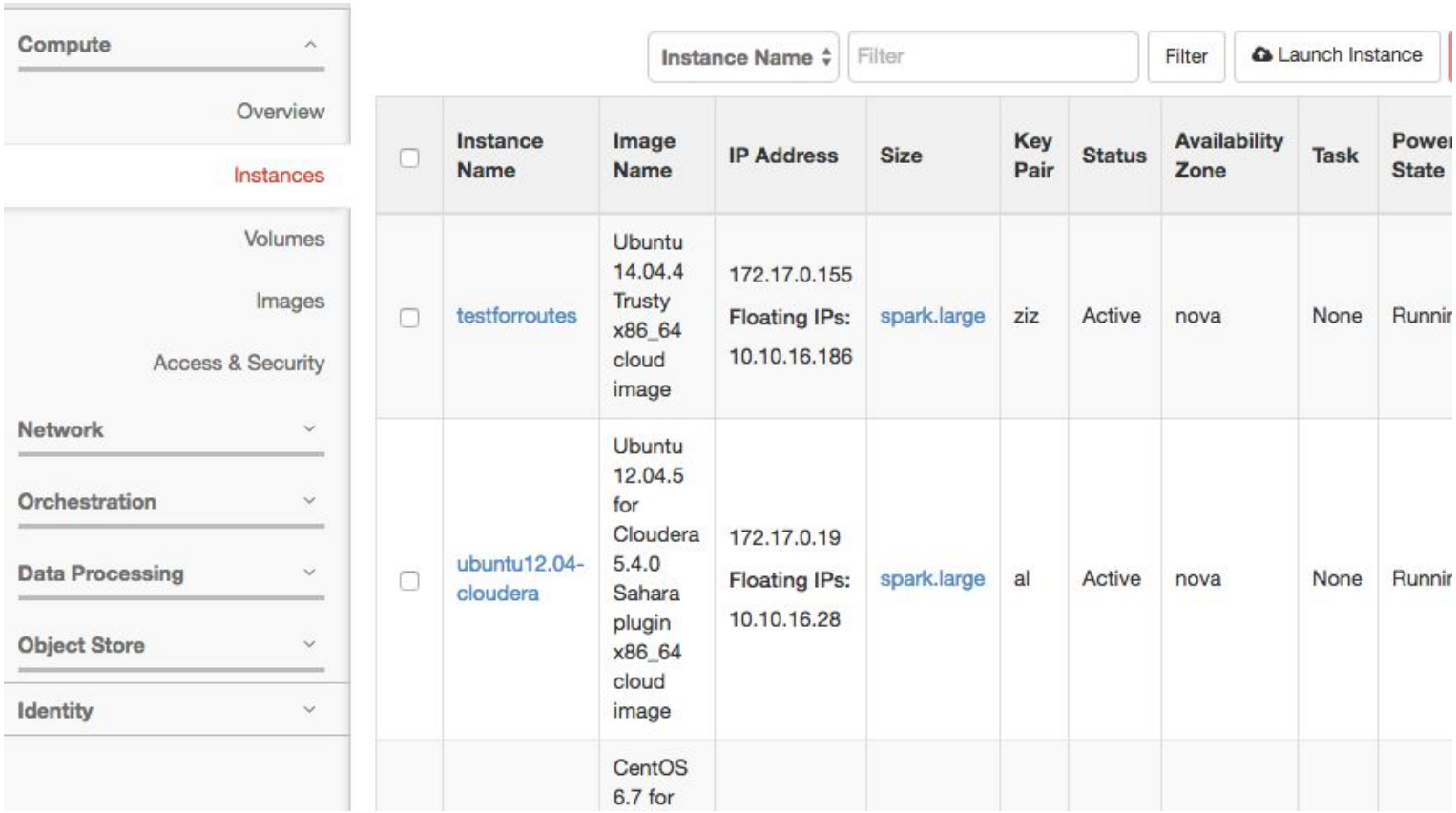


Доступ через API

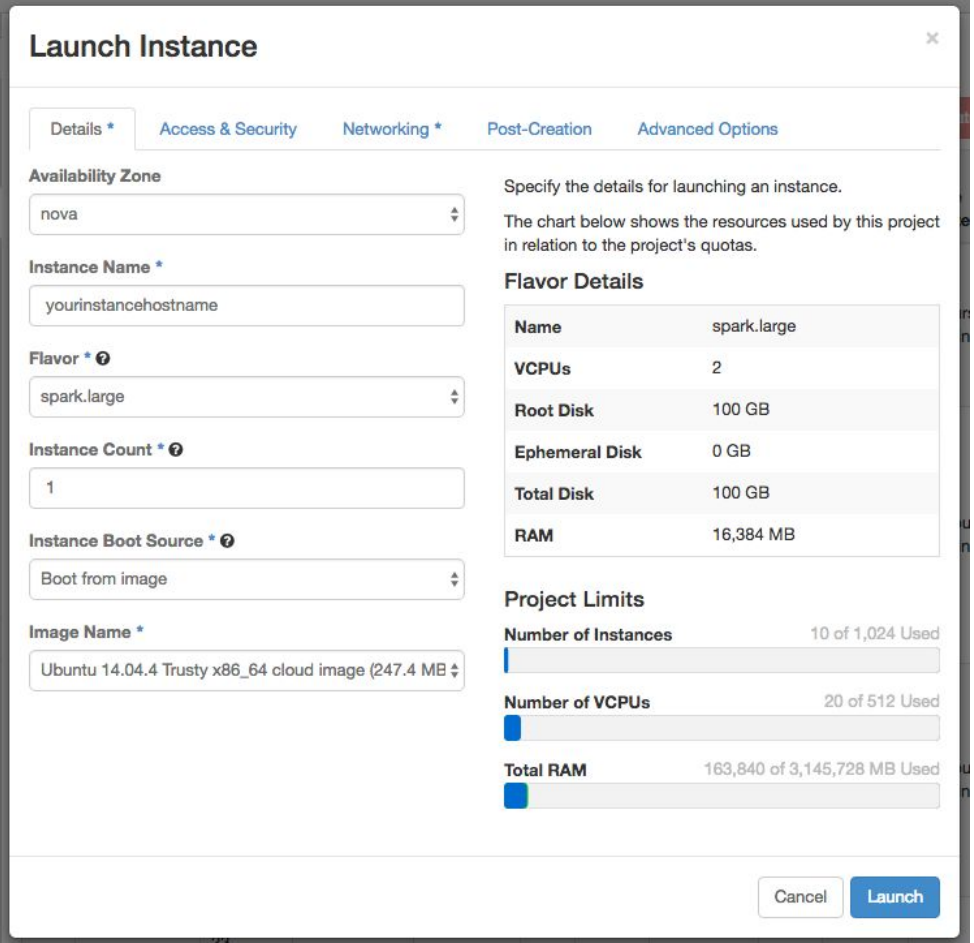
Для доступа через API или при помощи консольных клиентов, скачайте bash скрипт и сделайте `source <project-name>-openrc.sh` . Там прописываются переменные окружения для доступа к облаку.

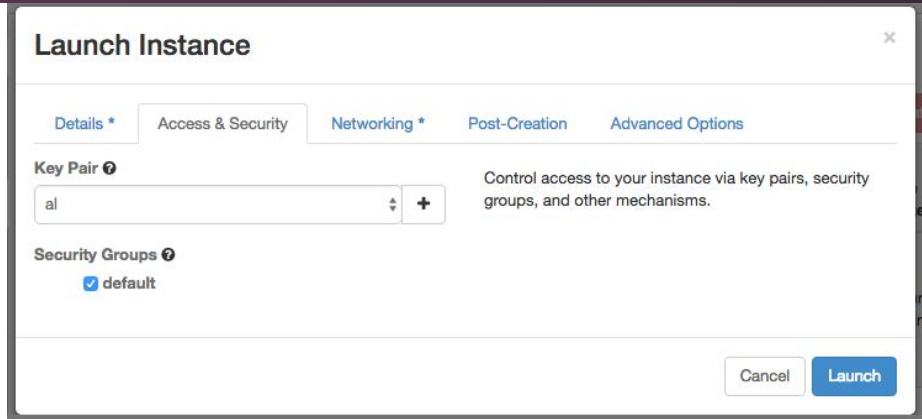


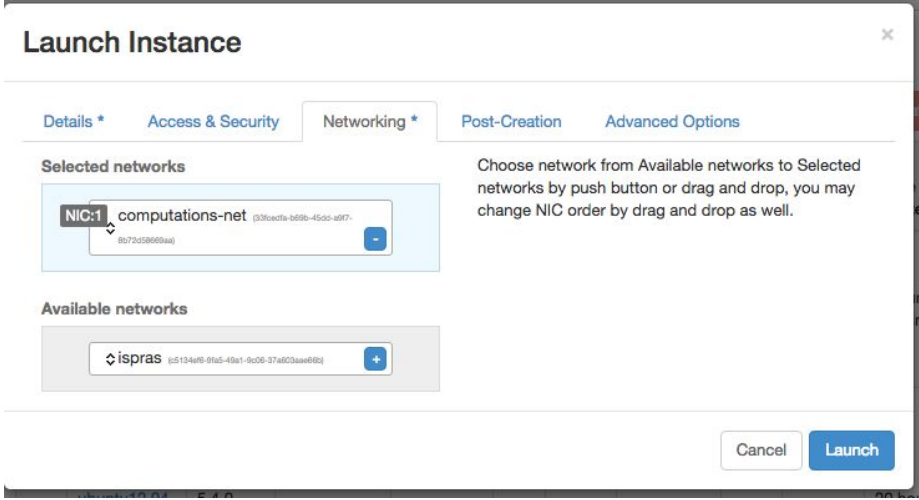
Запуск машины



Запуск машины







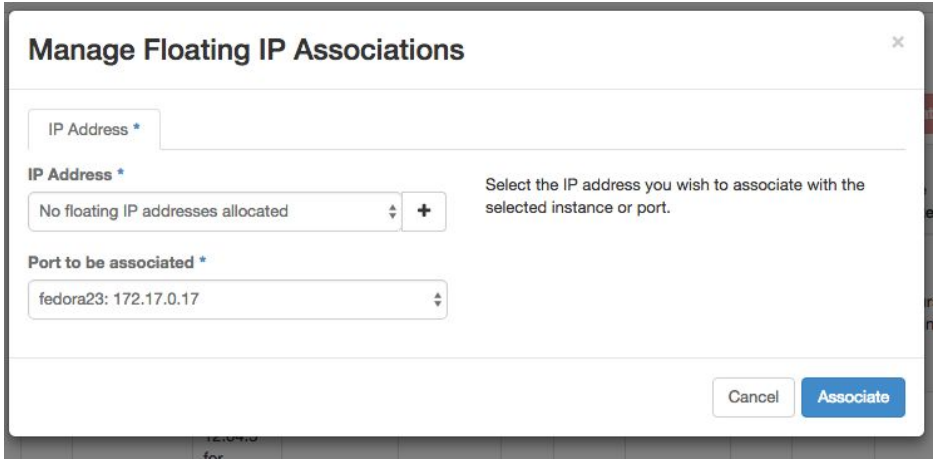
ИСП РАН

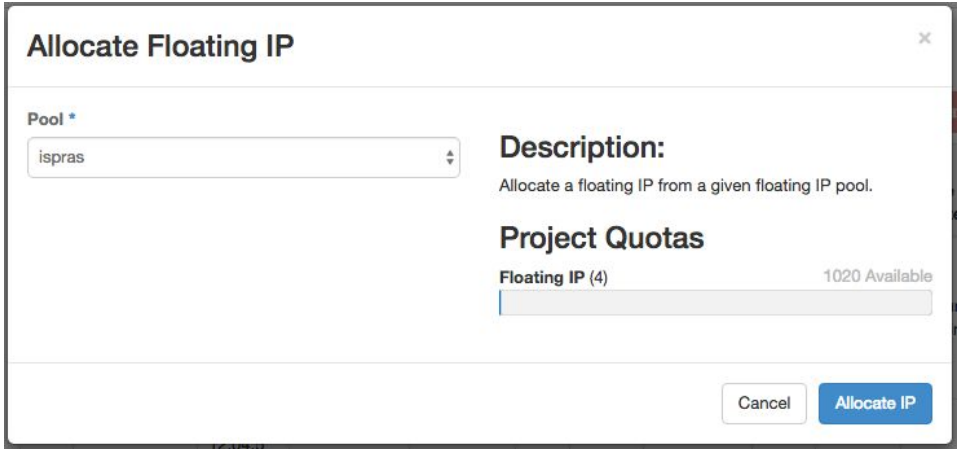
Запуск машины

Fedora 23 x86_64 cloud image	172.17.0.17	spark.large	al	Active	nova	None	Running	20 hours, 49 minutes	Create Snapshot

Associate Floating IP

Attach Interface



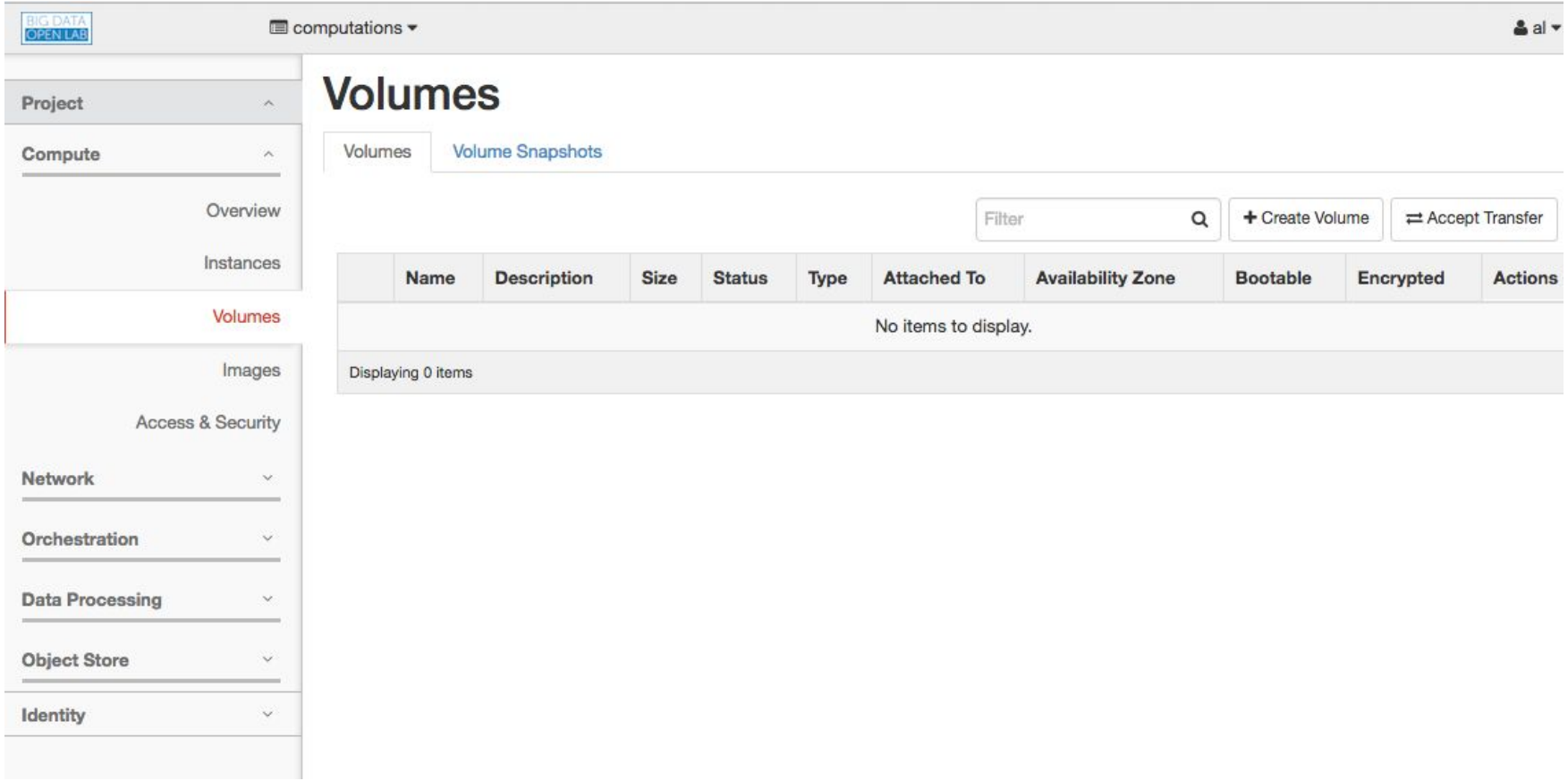


ИСП РАН

Про адреса и доступ

- Пожалуйста, в проектах computations и docker не аллоцируйте себе адреса из external_network без большой необходимости.
- External_network предназначена для внешних ресурсов.
- Адреса из сети ispras доступны внутри института и через VPN, адреса из external_network доступны отовсюду.
- У доступных сейчас образов логин совпадает с названием системы (ubuntu - ubuntu, centos - centos и т.д.)

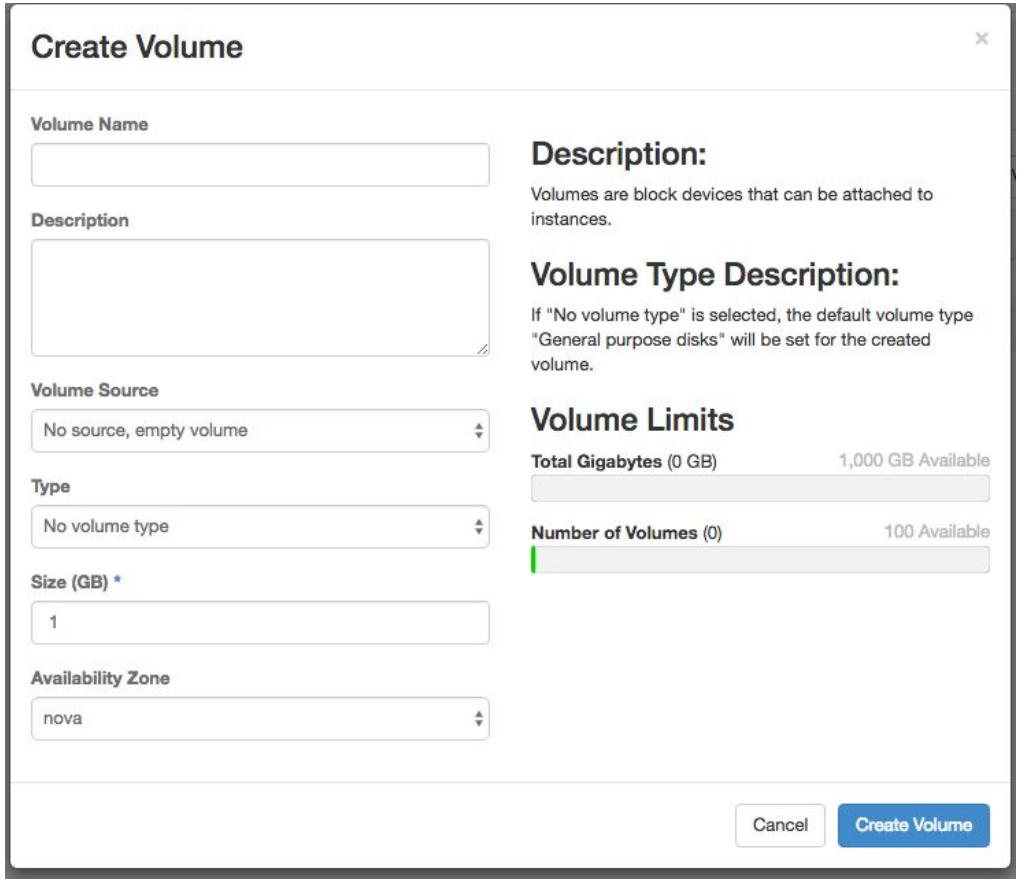
Создание надежного блочного устройства



ИСП РАН

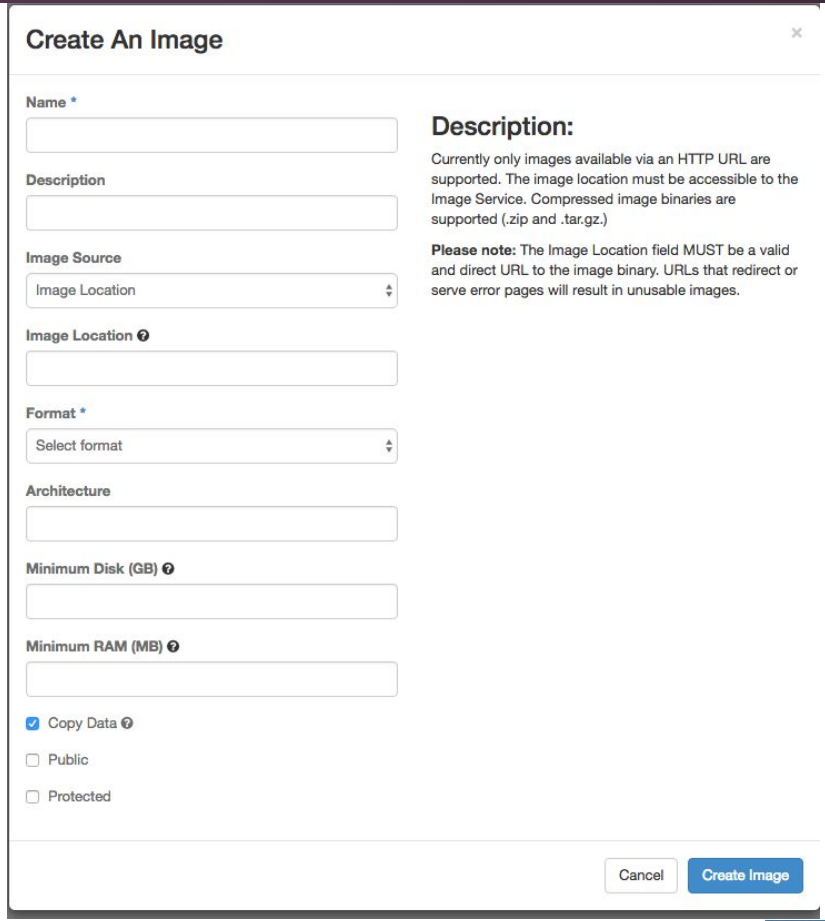
Создание надежного блочного устройства

- Можно создать пустой
- Можно создать из загрузочного образа
- Доступно две зоны:
- Большая - для обычного использования
- Маленькая - для работы с маленькими блоками

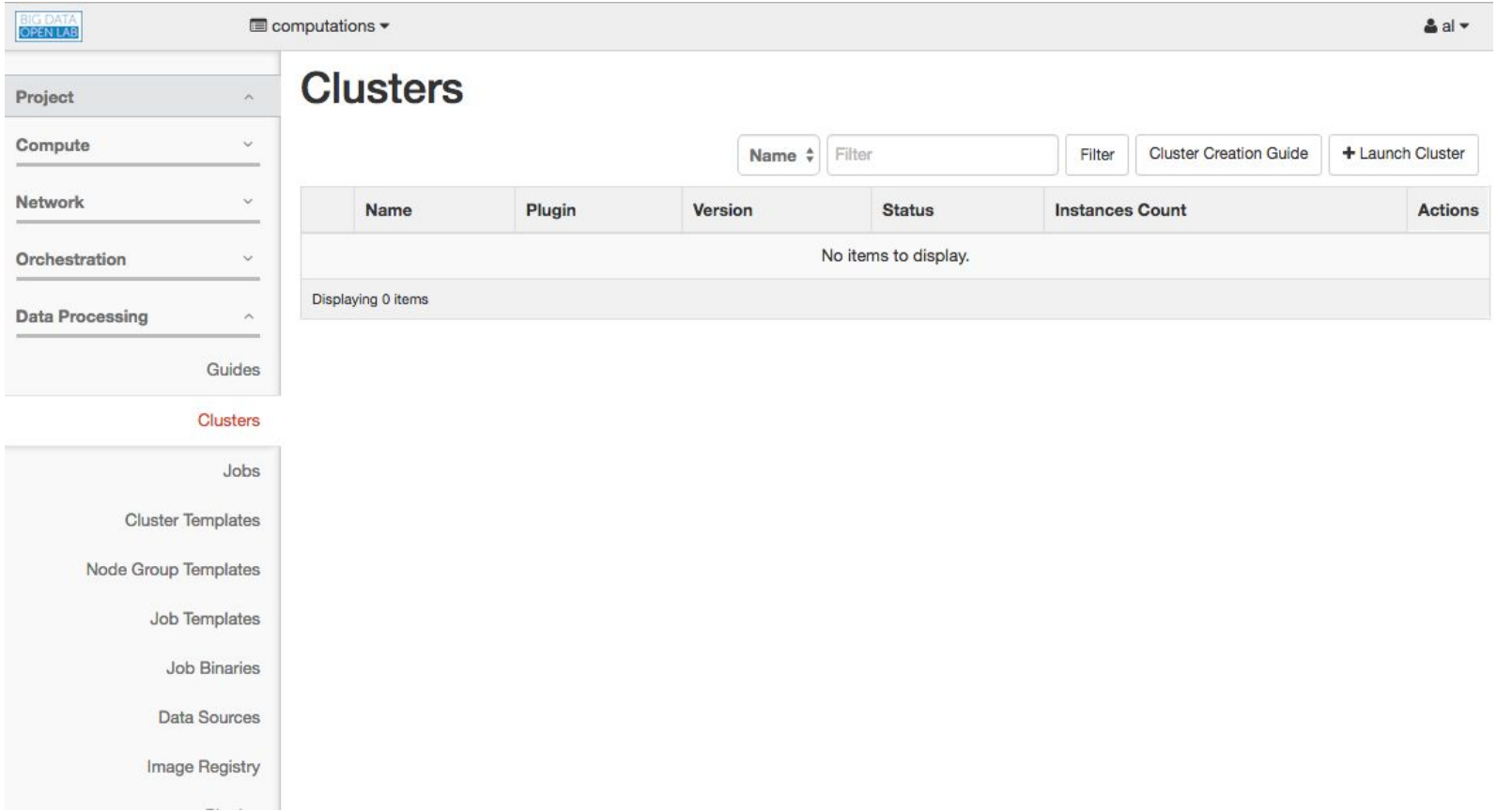


Создание загрузочного образа

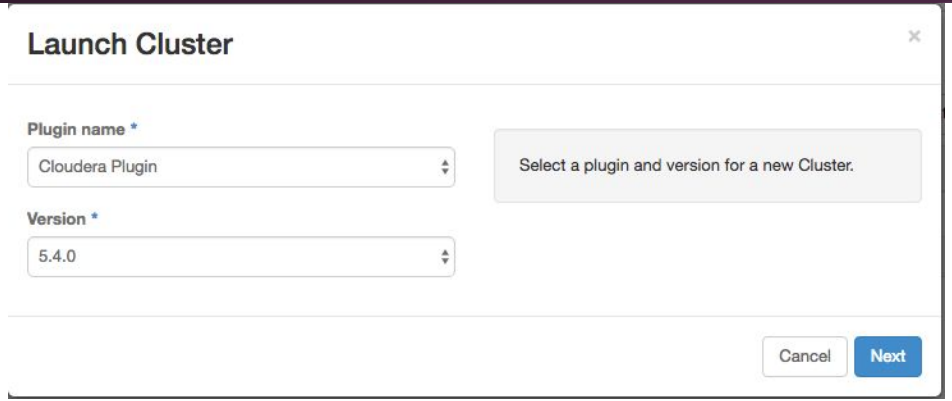
- На вход принимаются QCOW2, RAW и DOCKER образы (остальные не проверялись)
- В проекте computations нужно создавать образы, которые будут запускаться под KVM
- В проекте docker все иначе:
  - Вам нужен докер на локальной машине
  - Нужно построить образ и выполнить нечто подобное:
    - `sudo docker save docker/ubuntu-cloud-test | glance image-create --visibility=public --container-format=docker --disk-format=raw --name docker/ubuntu-cloud-test`
    - Имя образа в Glance обязано соответствовать имени образа в docker
- Не создавайте публичные образы без явной необходимости
- Protected позволяет защитить образ от удаления (пока галку кто-нибудь не снимет)

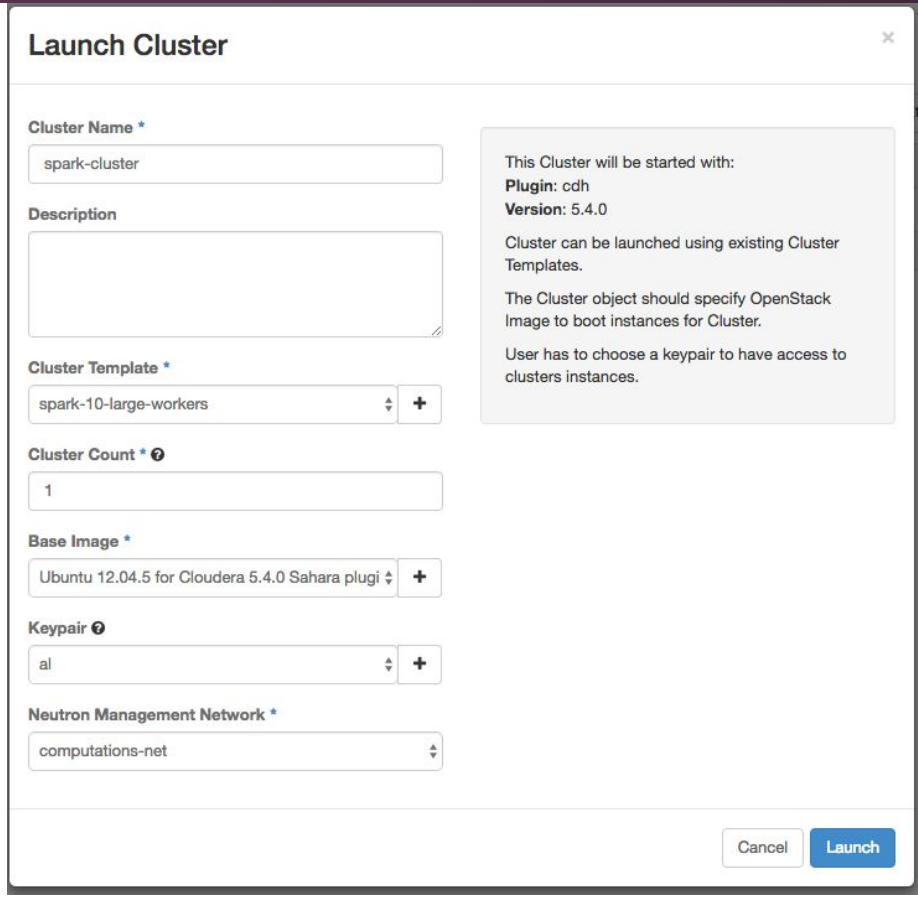


Big data processing (sahara)

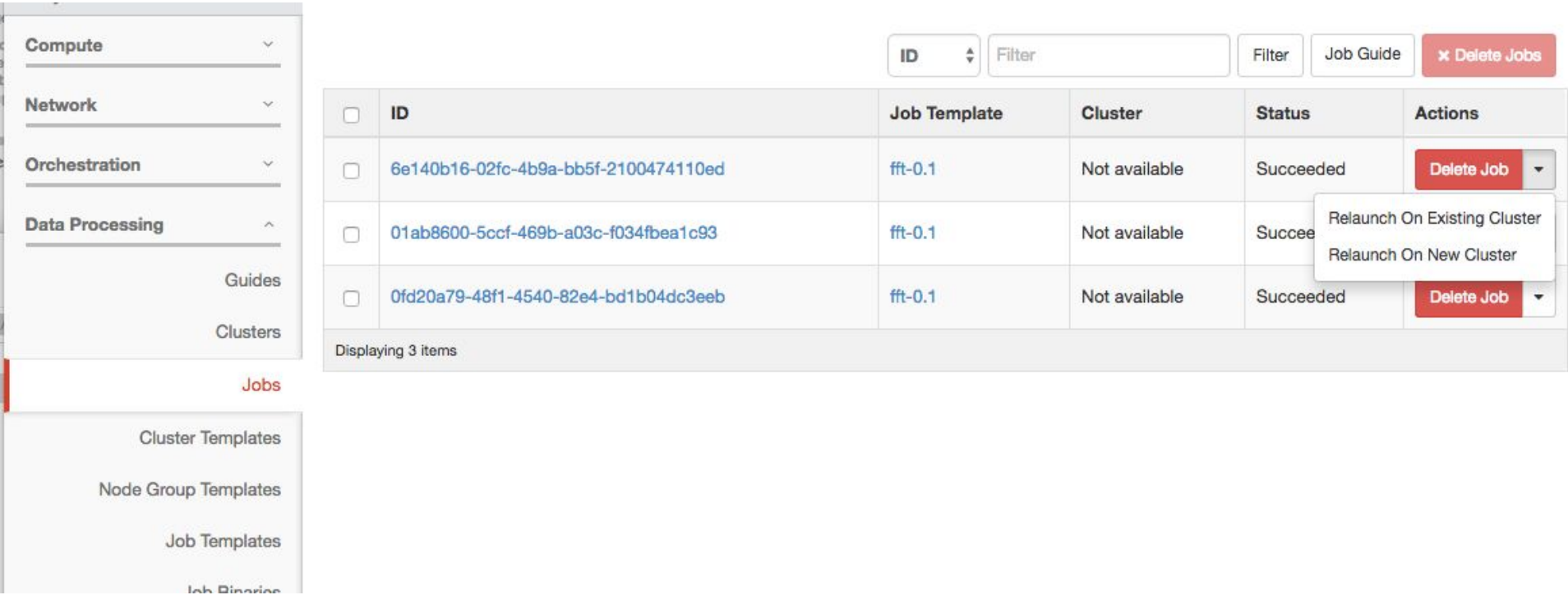


Big data processing (sahara)





Big data processing (sahara)



Странно работающие вещи

- Если виртуальная машина создана из Glance-образа, то ее snapshot попадет и в Glance, и в Cinder, причем из Glance запуститься не удастся
- Виртуальные машины не видят друг друга по именам (надеюсь, починим)
- Виртуальные машины вообще не знают ничего о своих внешних адресах (это by design)
- Докер-образы должны быть очень особенными (мне пока удалось построить один “условно рабочий”)
- Всем проектам доступны все внешние сети (особенность текущего релиза)

Побочные результаты

- Найден потенциальный баг в прошивках промышленных коммутаторов Dell (данные переданы)
- Найден серьезный баг в Openstack Cinder (починен и попал во все актуальные ветки)

Разделение на “проекты”

Мы сделали проекты для логического разбиения по потребностям:

- Computations — проект для кратковременных вычислений (посчитал-уничтожил). Работает под KVM.
- Docker — проект для экспериментов с Docker (с кучей ограничений)
- Outlanders — проект для чужаков
- Fancy — проект для демонстраций во внешний мир
- и несколько служебных проектов

Титры

Подбор железа и проектирование

Борисенко Олег

Закупка железа, спецификации, головная боль с формальностями

Калугин Миша

Эчина Даша

Самоваров Олег

Решение проблем с реформированием серверной, энергопотреблением и охлаждением

Андреев Олег

Самоваров Олег

Музыка Владимир Витальевич

Решение проблем с дефектным железом:

Шер Арсений

Борисенко Олег

Настройка всей системы:

Борисенко Олег

Алексиянц Саша

Шер Арсений

Жижченко Миша

Губенко Яша

Анализ производительности систем хранения

Шер Арсений

Борисенко Олег

ИСП РАН