

# Облако ИСП РАН

*наш маленький Amazon EC2*

# 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

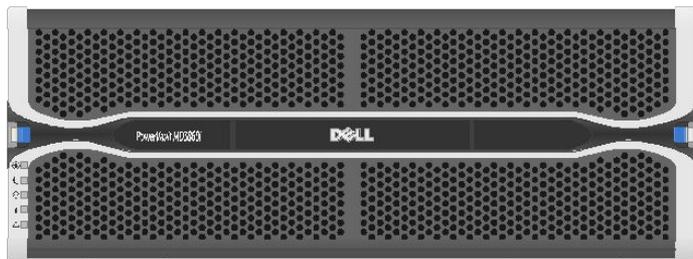
# Железо/хранение



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

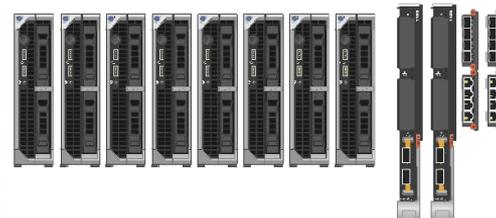
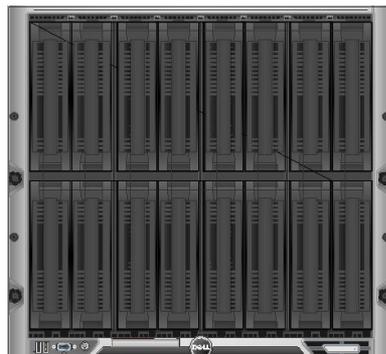
# Железо/вычисления



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  
2x 10Gbit/s SFP+  
1x 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

# Железо/сеть



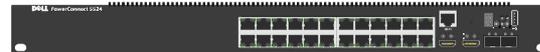
Dell N3024, 1U, 200W  
24x 1G bit/s  
2x 1G bit SFP  
2x 10G bit/s SFP+  
2x 10G bit/s Base-T (витая пара Cat6)



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

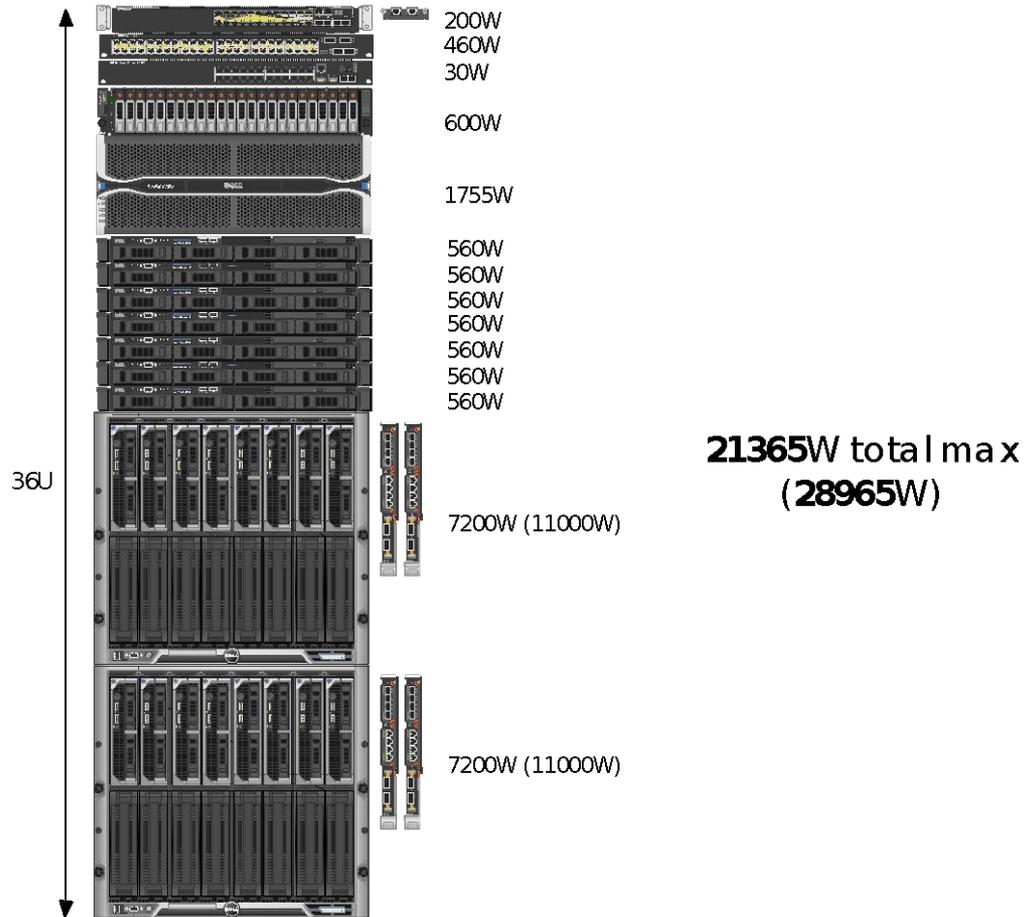


4x Dell MXL 10/40 Gbit/s:  
Встроенное подключение всех серверов  
в корзине  
2x 40Gbit/s  
4x 10Gbit/s SFP+  
4x 10Gbit/s Base-T (витая пара Cat6)



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

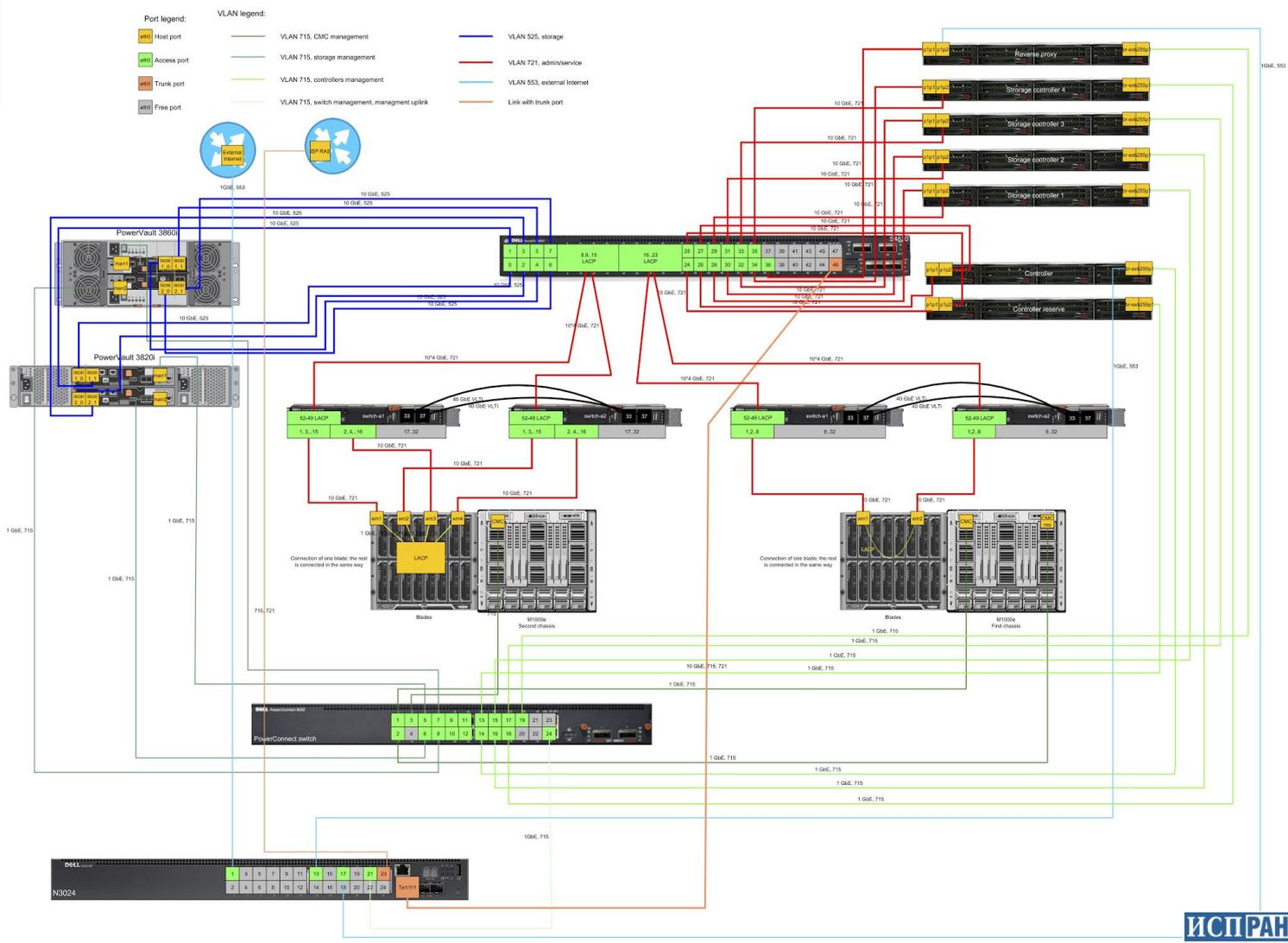
# Энергопотребление системы



# Сетевая часть

Векторная версия с увеличением тут:

<https://drive.google.com/file/d/0B4MHBK KyFpJfT3V3bmpyb2 9MUIE/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



## 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](mailto:borisenko@ispras.ru)

**АКТИВИРУЙ МЕНЯ**



## Log In

User Name

Password



Connect

# 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)			<a href="#">+ Create Image</a>	<a href="#">x Delete Images</a>
<input type="checkbox"/>	Image Name	Type	Status	Public	Protected	Format	Size	Actions
<input type="checkbox"/>	<a href="#">Ubuntu 12.04.5 for Cloudera 5.4.0 Sah...</a>	Image	Active	No	No	QCOW2	3.1 GB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">CentOS 6.7 for Cloudera 5.4.0 Sahara ...</a>	Image	Active	No	No	QCOW2	3.0 GB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">Ubuntu 14.04.4 Trusty x86_64 cloud image</a>	Image	Active	No	No	Raw	247.4 MB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">Ubuntu 12.04.5 Precise x86_64 cloud i...</a>	Image	Active	No	No	Raw	251.4 MB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">Debian 8.4 Jessie x86_64 cloud image</a>	Image	Active	No	No	QCOW2	458.8 MB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">Fedora 23 x86_64 cloud image</a>	Image	Active	No	No	QCOW2	223.5 MB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">CentOS 7 x86_64 cloud image</a>	Image	Active	No	No	QCOW2	872.9 MB	Launch Instance ▾
<input type="checkbox"/>	<a href="#">CentOS 6.7 x86_64 cloud image</a>	Image	Active	No	Yes	QCOW2	712.3 MB	Launch Instance ▾
Displaying 8 items								

# Запуск машины

Project ^

Compute ^

Overview

Instances

Volumes

Images

Access & Security

Network v

Orchestration v

Data Processing v

Object Store v

## Access & Security

Security Groups Key Pairs Floating IPs API Access

Filter



+ Create Key Pair

Import Key Pair

× Delete Key Pairs

<input type="checkbox"/>	Key Pair Name	Fingerprint	Actions
<input type="checkbox"/>	al	76:c2:ad:0a:1a:bd:3c:fb:f1:a8:bf:6d:5c:ef:f3:13	Delete Key Pair

Displaying 1 item

# 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.

**Access & Security**

Security Groups   Key Pairs   Floating IPs   **API Access**

[Download OpenStack RC File](#)   [+ View Credentials](#)

Service	Service Endpoint
Compute	http://cloud.ispras.ru:8774/v2/6aec0acc18aa44048f7da0a3144ef744
Network	http://cloud.ispras.ru:9696
Volumev2	http://cloud.ispras.ru:8776/v2/6aec0acc18aa44048f7da0a3144ef744
Image	http://cloud.ispras.ru:9292
Metering	http://cloud.ispras.ru:8777
Cloudformation	http://cloud.ispras.ru:8000/v1
Volume	http://cloud.ispras.ru:8776/v1/6aec0acc18aa44048f7da0a3144ef744
Orchestration	http://cloud.ispras.ru:8004/v1/6aec0acc18aa44048f7da0a3144ef744
Object Store	http://cloud.ispras.ru:8080/v1/AUTH_6aec0acc18aa44048f7da0a3144ef744
Data Processing	http://cloud.ispras.ru:8386/v1.1/6aec0acc18aa44048f7da0a3144ef744
Identity	https://cloud.ispras.ru:5000/v2.0

Displaying 11 items

# Запуск машины

Compute ^

Overview

Instances

Volumes

Images

Access & Security

Network ^

Orchestration ^

Data Processing ^

Object Store ^

Identity ^

Instance Name ^ Filter Filter  Launch Instance

<input type="checkbox"/>	Instance Name	Image Name	IP Address	Size	Key Pair	Status	Availability Zone	Task	Power State
<input type="checkbox"/>	<a href="#">testforroutes</a>	Ubuntu 14.04.4 Trusty x86_64 cloud image	172.17.0.155 Floating IPs: 10.10.16.186	<a href="#">spark.large</a>	ziz	Active	nova	None	Running
<input type="checkbox"/>	<a href="#">ubuntu12.04-cloudera</a>	Ubuntu 12.04.5 for Cloudera 5.4.0 Sahara plugin x86_64 cloud image	172.17.0.19 Floating IPs: 10.10.16.28	<a href="#">spark.large</a>	al	Active	nova	None	Running
		CentOS 6.7 for							

# Запуск машины

## Launch Instance

Details \* Access & Security Networking \* Post-Creation Advanced Options

### Availability Zone

nova

### Instance Name \*

yourinstancehostname

### Flavor \* ?

spark.large

### Instance Count \* ?

1

### Instance Boot Source \* ?

Boot from image

### Image Name \*

Ubuntu 14.04.4 Trusty x86\_64 cloud image (247.4 MB)

Specify the details for launching an instance.

The chart below shows the resources used by this project in relation to the project's quotas.

### Flavor Details

Name	spark.large
VCPUs	2
Root Disk	100 GB
Ephemeral Disk	0 GB
Total Disk	100 GB
RAM	16,384 MB

### Project Limits

Number of Instances 10 of 1,024 Used

Number of VCPUs 20 of 512 Used

Total RAM 163,840 of 3,145,728 MB Used

Cancel

Launch

## Launch Instance

Details \* Access & Security Networking \* Post-Creation Advanced Options

### Key Pair ?

al

Control access to your instance via key pairs, security groups, and other mechanisms.

### Security Groups ?

default

Cancel

Launch

## Launch Instance

Details \* Access & Security Networking \* Post-Creation Advanced Options

### Selected networks

NIC:1 computations-net (330c6cfa-b69b-45cc-af07-8b72c59890aa)

Choose network from Available networks to Selected networks by push button or drag and drop, you may change NIC order by drag and drop as well.

### Available networks

ispras (c5134af9-9fa5-49a1-9c06-37af03aae68b)

Cancel

Launch

# Запуск машины

Fedora 23 x86_64 cloud image	172.17.0.17	spark.large	al	Active	nova	None	Running	20 hours, 49 minutes	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Create Snapshot ▼</div> <div style="border: 1px solid #ccc; padding: 2px; display: inline-block; margin-top: 5px;">Associate Floating IP</div> <div style="border: 1px solid #ccc; padding: 2px; display: inline-block; margin-top: 5px;">Attach Interface</div>
------------------------------	-------------	-------------	----	--------	------	------	---------	----------------------	--

### Manage Floating IP Associations ×

**IP Address \***

No floating IP addresses allocated⌵+

Select the IP address you wish to associate with the selected instance or port.

**Port to be associated \***

fedora23: 172.17.0.17⌵

Cancel Associate

### Allocate Floating IP ×

**Pool \***

ispras⌵

**Description:**  
Allocate a floating IP from a given floating IP pool.

**Project Quotas**

Floating IP (4) 1020 Available

Cancel Allocate IP

# 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

The screenshot shows the 'Volumes' page in the OpenLab Big Data interface. The top navigation bar includes the 'BIG DATA OPEN LAB' logo, a 'computations' dropdown menu, and a user profile icon labeled 'al'. The left sidebar contains a navigation menu with categories like Project, Compute, Overview, Instances, Volumes (highlighted in red), Images, Access & Security, Network, Orchestration, Data Processing, Object Store, and Identity. The main content area is titled 'Volumes' and has two tabs: 'Volumes' (active) and 'Volume Snapshots'. Below the tabs is a search filter box and two buttons: '+ Create Volume' and '= Accept Transfer'. A table with columns for Name, Description, Size, Status, Type, Attached To, Availability Zone, Bootable, Encrypted, and Actions is present. The table is currently empty, displaying 'No items to display.' and 'Displaying 0 items'.

computations

BIG DATA  
OPEN LAB

al

## Volumes

Volumes Volume Snapshots

Filter

Name	Description	Size	Status	Type	Attached To	Availability Zone	Bootable	Encrypted	Actions
No items to display.									
Displaying 0 items									

# 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

## Create Volume

**Volume Name**

**Description**

**Volume Source**

No source, empty volume

**Type**

No volume type

**Size (GB) \***

1

**Availability Zone**

nova

**Description:**

Volumes are block devices that can be attached to instances.

**Volume Type Description:**

If "No volume type" is selected, the default volume type "General purpose disks" will be set for the created volume.

**Volume Limits**

Total Gigabytes (0 GB) 1,000 GB Available

Number of Volumes (0) 100 Available

Cancel Create Volume

# 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)

**Create An Image** ✕

Name \*

Description

Image Source

Image Location ⓘ

Format \*

Architecture

Minimum Disk (GB) ⓘ

Minimum RAM (MB) ⓘ

Copy Data ⓘ

Public

Protected

Cancel Create Image

**Description:**

Currently only images available via an HTTP URL are supported. The image location must be accessible to the Image Service. Compressed image binaries are supported (.zip and .tar.gz.)

**Please note:** The Image Location field MUST be a valid and direct URL to the image binary. URLs that redirect or serve error pages will result in unusable images.

# Big data processing (sahara)

BIG DATA OPEN LAB

computations ▾

al ▾

## Clusters

Name ▾ Filter Filter Cluster Creation Guide + Launch Cluster

Name	Plugin	Version	Status	Instances Count	Actions
No items to display.					
Displaying 0 items					

Project ^

Compute ▾

Network ▾

Orchestration ▾

Data Processing ^

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Data Sources

Image Registry

# Big data processing (sahara)

## Launch Cluster

Plugin name \*

Cloudera Plugin

Select a plugin and version for a new Cluster.

Version \*

5.4.0

Cancel

Next

## Launch Cluster

Cluster Name \*

spark-cluster

Description

Cluster Template \*

spark-10-large-workers

Cluster Count \* ?

1

Base Image \*

Ubuntu 12.04.5 for Cloudera 5.4.0 Sahara plugi

Keypair ?

al

Neutron Management Network \*

computations-net

This Cluster will be started with:

**Plugin:** cdh

**Version:** 5.4.0

Cluster can be launched using existing Cluster Templates.

The Cluster object should specify OpenStack Image to boot instances for Cluster.

User has to choose a keypair to have access to clusters instances.

Cancel

Launch

# Big data processing (sahara)

- Compute
- Network
- Orchestration
- Data Processing**

Guides

Clusters

**Jobs**

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

ID Filter Filter Job Guide **✕ Delete Jobs**

<input type="checkbox"/>	ID	Job Template	Cluster	Status	Actions
<input type="checkbox"/>	<a href="#">6e140b16-02fc-4b9a-bb5f-2100474110ed</a>	fft-0.1	Not available	Succeeded	<b>Delete Job</b>
<input type="checkbox"/>	<a href="#">01ab8600-5ccf-469b-a03c-f034fbea1c93</a>	fft-0.1	Not available	Succeeded	Relaunch On Existing Cluster Relaunch On New Cluster
<input type="checkbox"/>	<a href="#">0fd20a79-48f1-4540-82e4-bd1b04dc3eeb</a>	fft-0.1	Not available	Succeeded	<b>Delete Job</b>

Displaying 3 items

# 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