

Открытый облачный сервис OpenStack и поддержка инструментария Spark средствами Sahara

Алексиянц Александр, ИСП РАН

Основные возможности облачных сред

- Выделение виртуальных машин с заданными характеристиками по запросу
- Построение виртуальных сетевых инфраструктур
- Виртуализация блочного хранения
- Объектные хранилища
- Встроенные системы оркестрации ресурсов
- Все это можно использовать при помощи простого REST API

Основные особенности

В чем преимущества для пользователей:

- Не нужно иметь собственный парк железа; плата только за использованные ресурсы
- Возможность адаптироваться под нагрузку
- Виртуализация как вычислительной части, так сетей и хранения
- Низкий порог необходимых знаний для использования
- Можно выбирать ресурсы под задачу, а не использовать “что есть”

В чем преимущества для провайдеров облачных услуг:

- Дешевле в плане электроэнергии
- Можно продавать более сложные вещи (PaaS) за больше денег, чем “просто хостинг”

Самые известные из коммерческих – Amazon EC2 и Microsoft Azure

Самые известных из открытых – Openstack и Eucalyptus

Amazon

Самый полный
набор решений:

Compute

-  **EC2**
Virtual Servers in the Cloud
-  **EC2 Container Service**
Run and Manage Docker Containers
-  **Elastic Beanstalk**
Run and Manage Web Apps
-  **Lambda**
Run Code in Response to Events

Storage & Content Delivery

-  **S3**
Scalable Storage in the Cloud
-  **CloudFront**
Global Content Delivery Network
-  **Elastic File System PREVIEW**
Fully Managed File System for EC2
-  **Glacier**
Archive Storage in the Cloud
-  **Import/Export Snowball**
Large Scale Data Transport
-  **Storage Gateway**
Hybrid Storage Integration

Database

-  **RDS**
Managed Relational Database Service
-  **DynamoDB**
Managed NoSQL Database
-  **ElastiCache**
In-Memory Cache
-  **Redshift**
Fast, Simple, Cost-Effective Data Warehousing
-  **DMS**
Managed Database Migration Service

Networking

-  **VPC**
Isolated Cloud Resources
-  **Direct Connect**
Dedicated Network Connection to AWS
-  **Route 53**
Scalable DNS and Domain Name Registration

Developer Tools

-  **CodeCommit**
Store Code in Private Git Repositories
-  **CodeDeploy**
Automate Code Deployments
-  **CodePipeline**
Release Software using Continuous Delivery

Management Tools

-  **CloudWatch**
Monitor Resources and Applications
-  **CloudFormation**
Create and Manage Resources with Templates
-  **CloudTrail**
Track User Activity and API Usage
-  **Config**
Track Resource Inventory and Changes
-  **OpsWorks**
Automate Operations with Chef
-  **Service Catalog**
Create and Use Standardized Products
-  **Trusted Advisor**
Optimize Performance and Security

Security & Identity

-  **Identity & Access Management**
Manage User Access and Encryption Keys
-  **Directory Service**
Host and Manage Active Directory
-  **Inspector PREVIEW**
Analyze Application Security
-  **WAF**
Filter Malicious Web Traffic
-  **Certificate Manager**
Provision, Manage, and Deploy SSL/TLS Certificates

Analytics

-  **EMR**
Managed Hadoop Framework
-  **Data Pipeline**
Orchestration for Data-Driven Workflows
-  **Elasticsearch Service**
Run and Scale Elasticsearch Clusters
-  **Kinesis**
Work with Real-Time Streaming Data
-  **Machine Learning**
Build Smart Applications Quickly and Easily

Internet of Things

-  **AWS IoT**
Connect Devices to the Cloud

Game Development

-  **GameLift**
Deploy and Scale Session-based Multiplayer Games

Mobile Services

-  **Mobile Hub**
Build, Test, and Monitor Mobile Apps
-  **Cognito**
User Identity and App Data Synchronization
-  **Device Farm**
Test Android, FireOS, and iOS Apps on Real Devices in the Cloud
-  **Mobile Analytics**
Collect, View and Export App Analytics
-  **SNS**
Push Notification Service

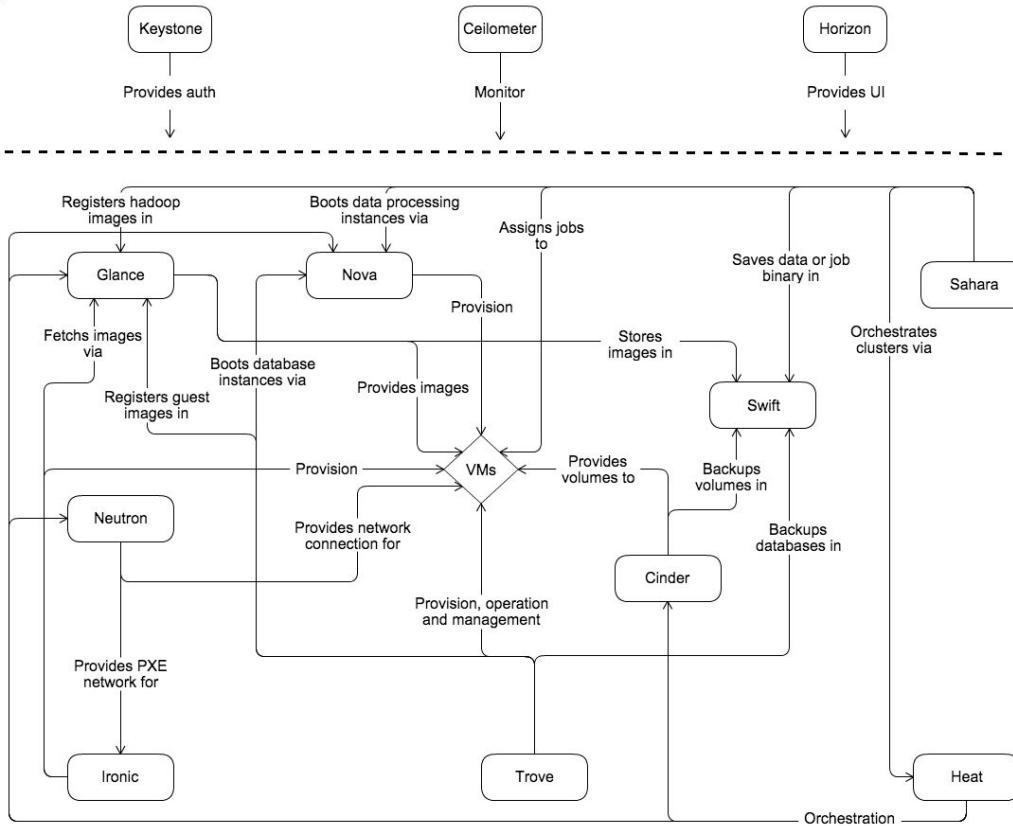
Application Services

-  **API Gateway**
Build, Deploy and Manage APIs
-  **AppStream**
Low Latency Application Streaming
-  **CloudSearch**
Managed Search Service
-  **Elastic Transcoder**
Easy-to-Use Scalable Media Transcoding
-  **SES**
Email Sending and Receiving Service
-  **SQS**
Message Queue Service
-  **SWF**
Workflow Service for Coordinating Application Components

Enterprise Applications

-  **WorkSpaces**
Desktops in the Cloud
-  **WorkDocs**
Secure Enterprise Storage and Sharing Service
-  **WorkMail**
Secure Email and Calendaring Service

Openstack, компоненты

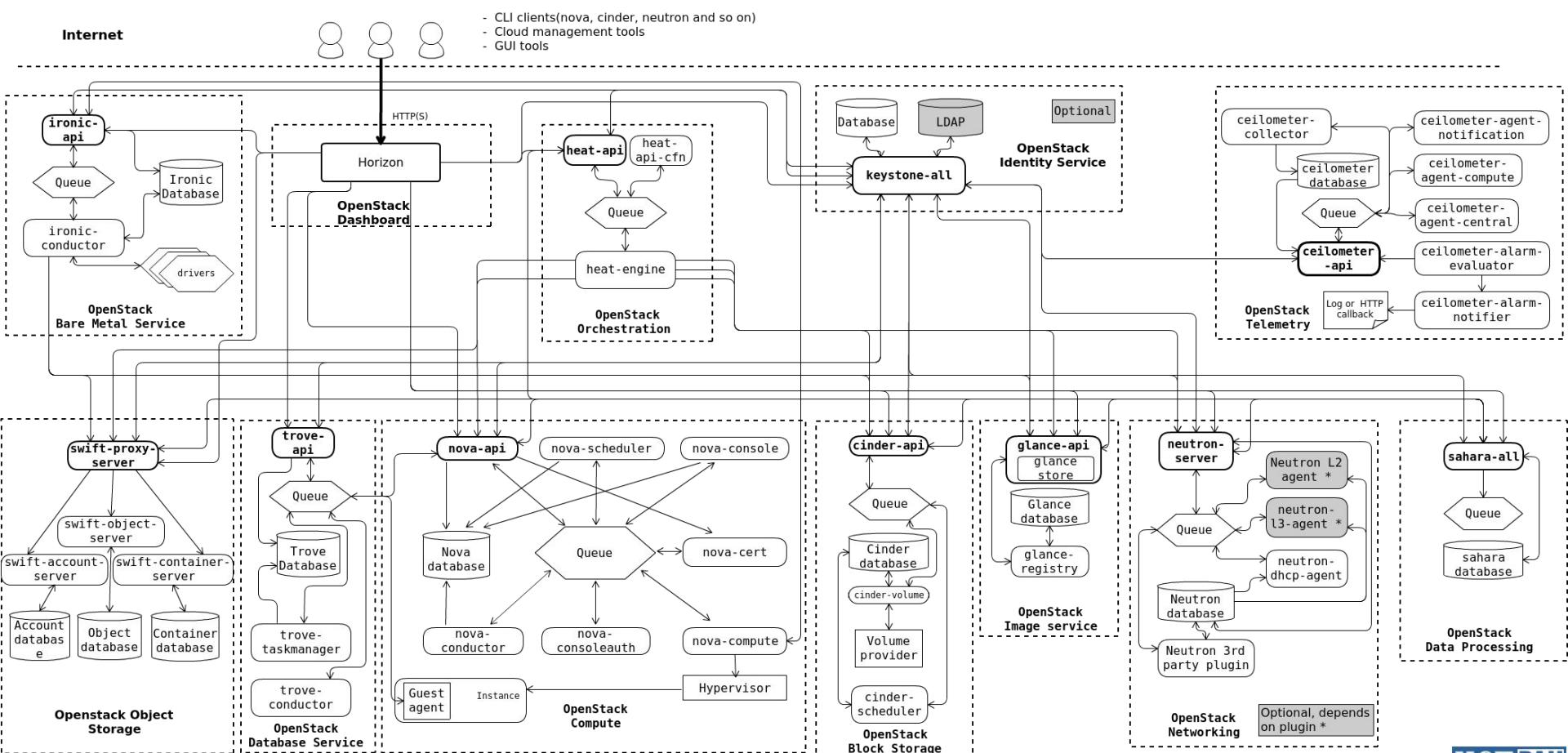


Openstack состоит из т.н. “проектов”. В реальности их больше, чем на этой схеме, основные из них:

- Keystone
- Nova
- Horizon
- Neutron
- Cinder
- Glance
- Swift
- Heat
- Sahara
- Ceilometer

Минимальный необходимый набор – Keystone, Glance и Nova.

Openstack, полная архитектура

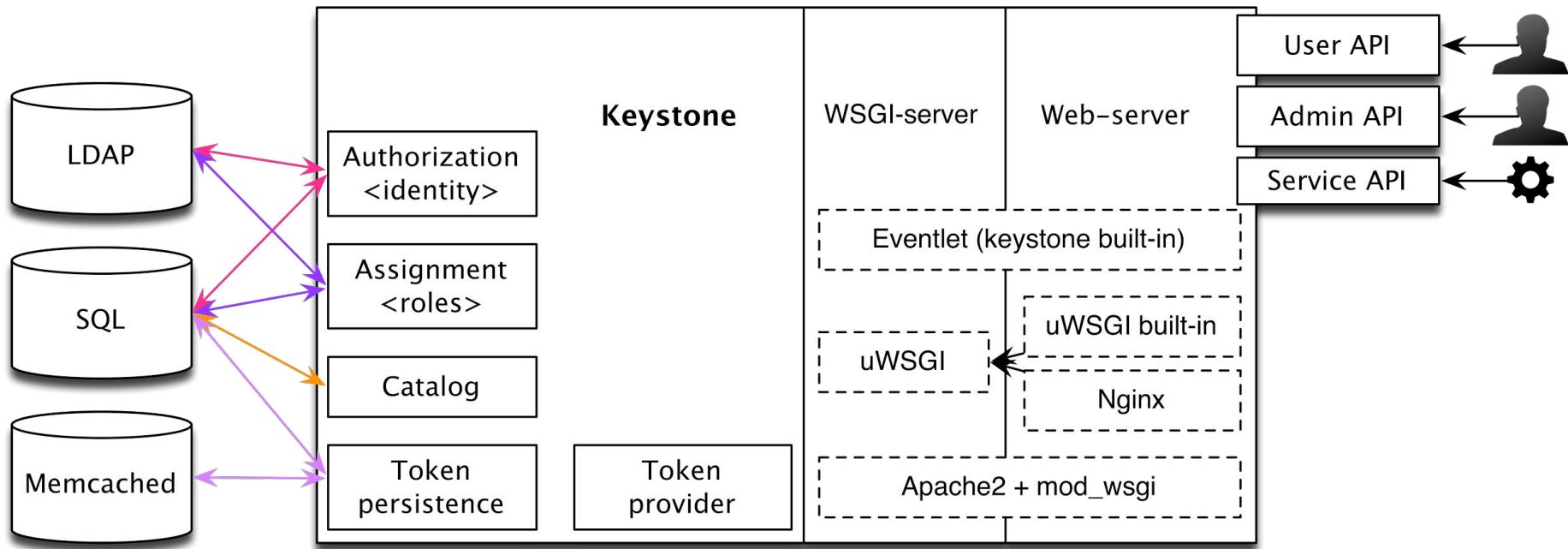


Keystone

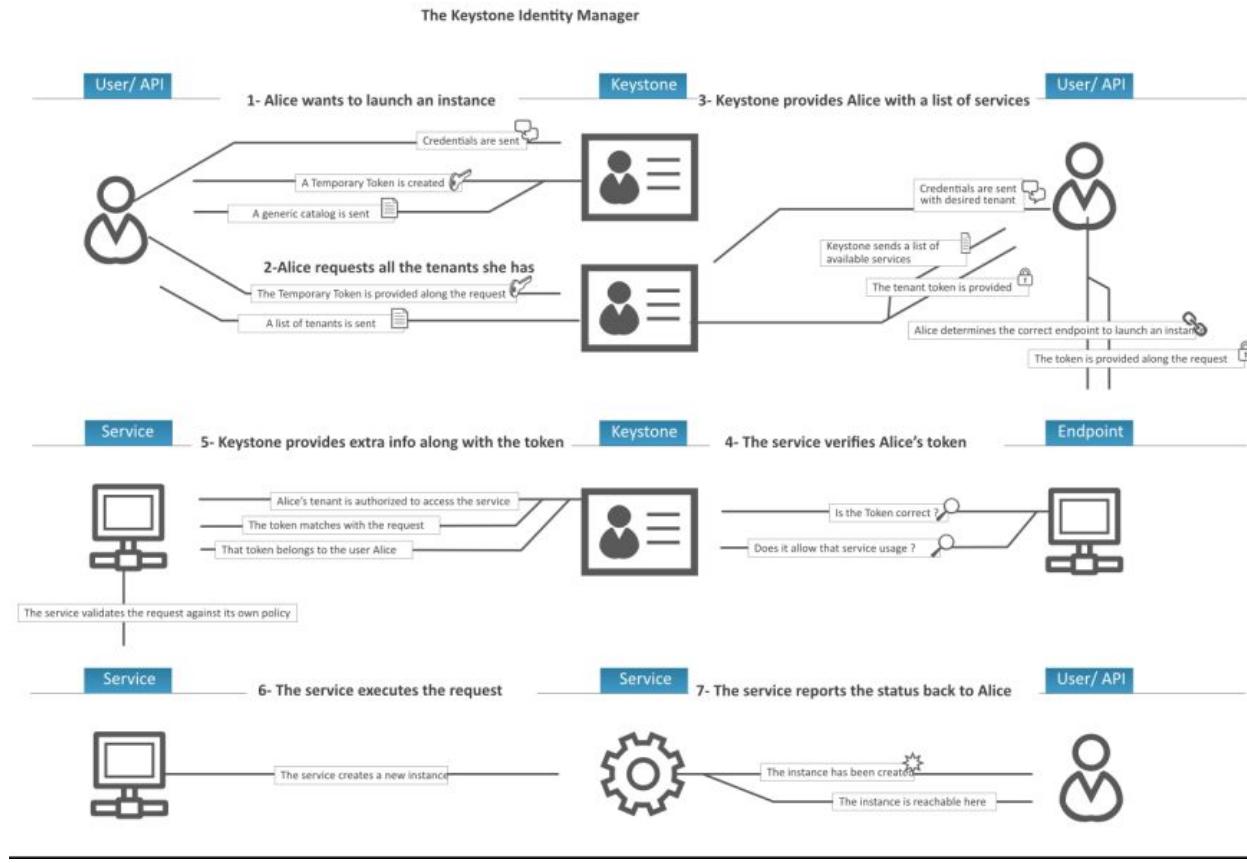
Центральный сервис аутентификации и авторизации.

- Обслуживает и пользователей, и сами сервисы
- Работает на основе токенов
- В качестве бэкенда может использовать другие сервисы аутентификации
- Это обычное WSGI-приложение

Keystone



Keystone



Keystone

Project

Compute

Overview

Instances

Volumes

Images

Access & Security

Network

Orchestration

Data Processing

Object Store

Admin

Identity

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/d4b175966deb46ac8280fcfc149a4457
Network	http://cloud.ispras.ru:9696
Volumev2	http://cloud.ispras.ru:8776/v2/d4b175966deb46ac8280fcfc149a4457
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/d4b175966deb46ac8280fcfc149a4457
Orchestration	http://cloud.ispras.ru:8004/v1/d4b175966deb46ac8280fcfc149a4457
Object Store	http://cloud.ispras.ru:8080/v1/AUTH_d4b175966deb46ac8280fcfc149a4457
Data Processing	http://cloud.ispras.ru:8386/v1.1/d4b175966deb46ac8280fcfc149a4457
Identity	https://cloud.ispras.ru:5000/v2.0

Displaying 11 items

ИСПРАН

Keystone

Project

Admin

Identity

Projects

Users

Displaying 5 items

Projects

Edit Project

Project Information * Project Members Quota *

Name
admin
service
demo
computations
outlanders

All Users Filter Q

nova	+
neutron	+
glance	+
cinder	+
sahara	+
heat	+
ceilometer	+
ceph	+
linuxcontest	+
clouddemo	+

Project Members Filter Q

admin member ▾	-
shaman member ▾	-
ai member ▾	-
fomin member ▾	-

Enabled Actions

80fcfc149a4457	Yes	Manage Members ▾
e6155b1c8f491d	Yes	Manage Members ▾
03ebd25f76e81	Yes	Manage Members ▾
da0a3144ef744	Yes	Manage Members ▾
83ac261a7b136	Yes	Manage Members ▾

Keystone

Project

Compute

Overview

Instances

Volumes

Images

Access & Security

Network

Orchestration

Data Processing

Object Store

Admin

Identity

Access & Security

Security Groups Key Pairs Floating IPs API Access

Filter Search icon + Create Key Pair Import Key Pair Delete Key Pairs

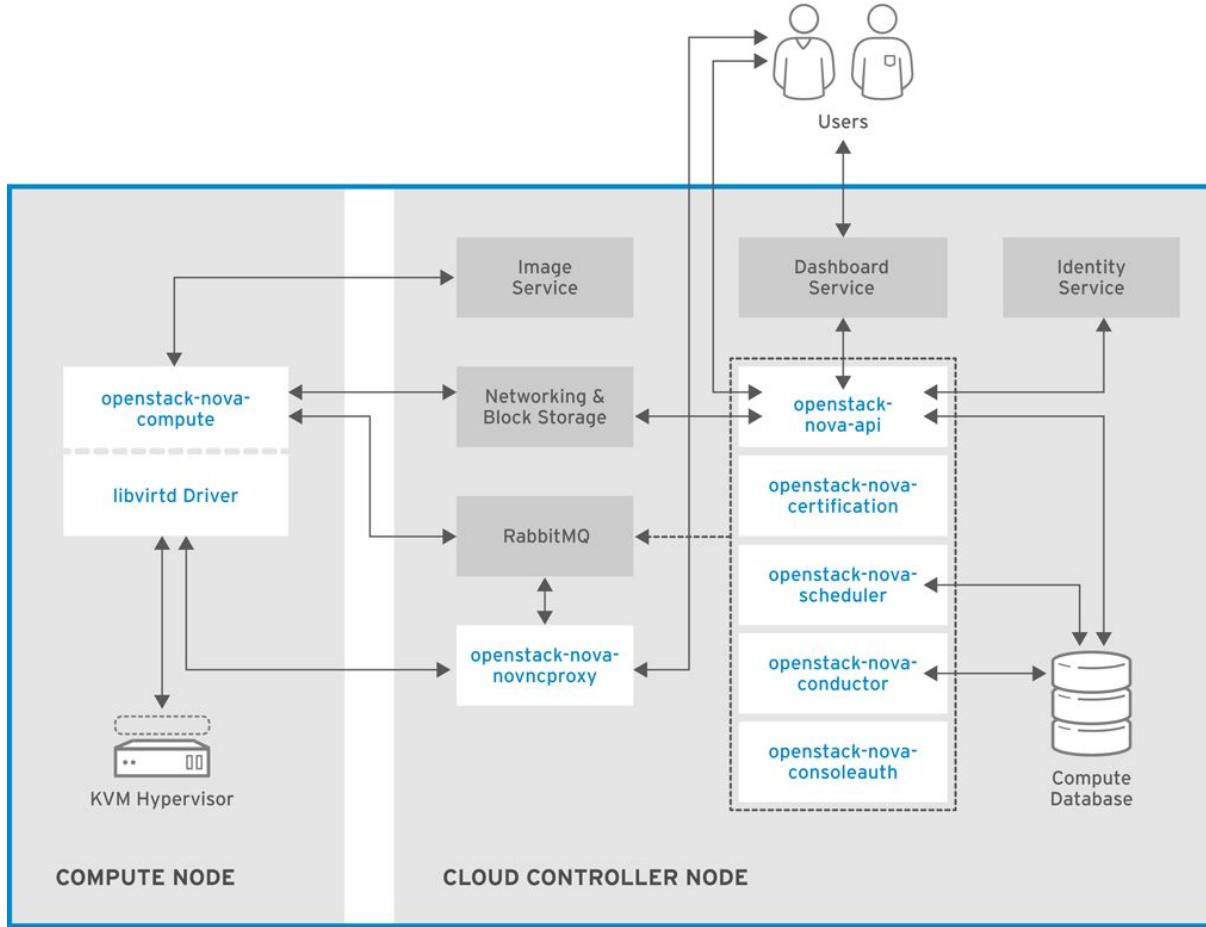
<input type="checkbox"/>	Key Pair Name	Fingerprint	Actions
<input type="checkbox"/>	al	76:c2:ad:0a:1a:bd:3c:fb:f1:a8:bf:6d:5c:ef:f3:13	Delete Key Pair
<input type="checkbox"/>	alexander	dd:88:7b:5b:9e:e8:dd:4e:63:2a:95:be:46:38:c3:63	Delete Key Pair
<input type="checkbox"/>	borisenko	ff:68:b8:02:2c:e4:d2:d9:0f:0e:03:43:34:7f:12:a4	Delete Key Pair
<input type="checkbox"/>	linux_contest	04:e4:ba:e4:43:82:fb:35:3d:5a:86:a3:67:41:78:3c	Delete Key Pair

Displaying 4 items

Nova

- Сервис предоставляет виртуальные машины.
- Умеет работать с разными гипервизорами.
- Умеет обслуживать Plain-сети (без Neutron).
- Отвечает за выделение физических ресурсов.

Nova



Nova

openstack

Project

Compute

Overview

Instances

Volumes

Images

Access & Security

Network

Orchestration

Data Processing

Object Store

Admin

Identity

Instances

Displaying 2 items

Instance Name

testintra

test-reso

Availability Zone

nova

Instance Name *

Instance Count *

1

Flavor * ?

func.testing

Instance Boot Source * ?

Select source

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

Launch Instance

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	func.testing
VCPUs	2
Root Disk	3 GB
Ephemeral Disk	3 GB
Total Disk	6 GB
RAM	256 MB

Project Limits

Number of Instances	2 of 1,000 Used
Number of VCPUs	3 of 512 Used
Total RAM	16,896 of 4,096,000 MB Used

Cancel Launch

Terminate Instances More Actions

Power State

Time since created

Actions

On State

2 days, 23 hours

Edit Instance

Shut Down

5 days, 2 hours

Start Instance

Nova

openstack admin

Project Compute Instances Volumes Images Access & Security Network Orchestration Data Processing Object Store Admin Identity

Instances

Displaying 2 items

Launch Instance

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

Key Pair ? al

Security Groups ? default

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

Cancel Launch

Instances Terminate Instances More Actions

Power State	Time since created	Actions
State	2 days, 23 hours	Edit Instance
Put Down	5 days, 2 hours	Start Instance

Nova

openstack admin

Project Compute Network Orchestration Data Processing Object Store Admin Identity

Instances

Overview Instances Volumes Images Access & Security

Displaying 2 items

Launch Instance

Details * Access & Security Networking *

Selected networks

- NIC:1 ispras (b6134ef8-8fa5-49a1-9c06-37af03aae68c)
- NIC:2 test-net (b70a639c-8467-4cf9-abf1-cf2c298fa515)

Available networks

- external_network (f62bd7af-bfc7-4a8c-8c0d-aa0be7a6a70)

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.

Power State: Running Time since created: 2 days, 23 hours Actions: Edit Instance

Power State: Shut Down Time since created: 5 days, 2 hours Actions: Start Instance

Cancel Launch

Horizon

Веб-интерфейс для Openstack.

- Веб-приложение на Django.
- Позиционируется как “фреймворк для построения веб-интерфейсов к Openstack”
- Модульный.
- Использует нативные библиотеки для Openstack на Python для коммуникации с Openstack.

Neutron

Сервис виртуальных сетей.

- Предоставляет изоляцию служебных и публичных сетей внутри Openstack (при помощи Vlan, VXLAN, GRE)
- Следит за Security группами (по сути - фаерволл)
- Позволяет программно задавать сетевую топологию виртуальных кластеров.
- Предоставляет публичные IP-адреса виртуальным машинам (реализовано как floating IP)

Neutron

Network Topology

Resize the canvas by scrolling up/down with your mouse/trackpad on the topology. Pan around the canvas by clicking and dragging the space behind the topology.

Toggle labels

Toggle Network Collapse

Launch Instance

Create Network

Create Router

Network Topology

Networks

Routers

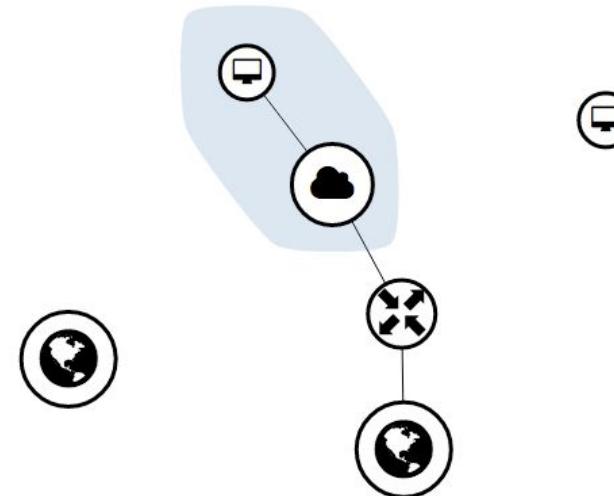
Orchestration

Data Processing

Object Store

Admin

Identity



Neutron

Network

Network Topology

Networks

Routers

Orchestration

Data Processing

Object Store

Admin

Identity

Network Overview

Name	test-net
ID	b70a639c-8467-4d9f-afb1-cf2c28fbba515
Project ID	d4b175966deb46ac8280fcfc149a4457
Status	Active
Admin State	UP
Shared	No
External Network	No
MTU	Unknown
Provider Network	Network Type: vxlan Physical Network: - Segmentation ID: 1016

Subnets

<input type="checkbox"/>	Name	Network Address	IP Version	Gateway IP	Actions
<input type="checkbox"/>	test-subnet	192.168.100.0/24	IPv4	192.168.100.1	<button>Edit Subnet</button> <button>▼</button>

Displaying 1 item

Ports

Name	Fixed IPs	Attached Device	Status	Admin State	Actions
(1cc650b7-bc95)	192.168.100.3	compute:nova	Down	UP	<button>Edit Port</button>
(492365de-06ae)	192.168.100.4	compute:nova	Down	UP	<button>Edit Port</button>

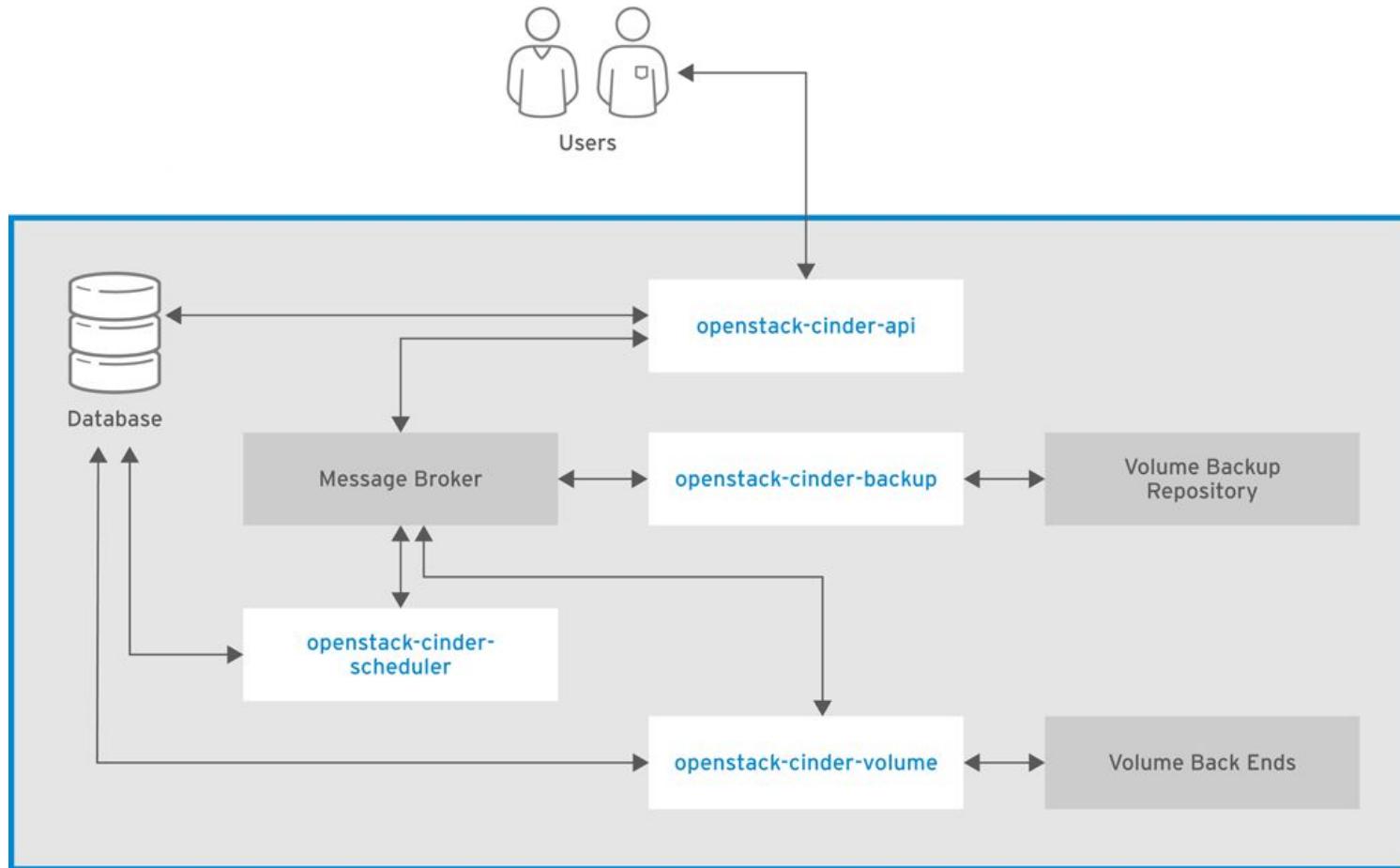
ИСПРАН

Cinder

Сервис предоставления блочных устройств виртуальным машинам.

- Умеет делать снимки образов.
- Отдает все по iSCSI.
- Поддерживает множественные бэкенды для хранения.

Cinder



Cinder

openstack admin

Project

Compute

Overview

Instances

Volumes

Images

Access & Security

Network

Orchestration

Data Processing

Object Store

Admin

Identity

admin

Volumes

Volumes Volume Snapshots

Filter + Create Volume

<input type="checkbox"/>	Name	Description	Size	Status	Type	Attached To	Availability Zone	Bootable	Encrypted	Actions
<input type="checkbox"/>	ubuntu-test	-	10GB	Available	General purpose disks		nova	Yes	No	<input type="button" value="Edit Volume"/>

Displaying 1 item

Cinder

openstack admin admin

Project Compute Overview Instances Volumes Images Access & Security Network Orchestration Data Processing Object Store Admin Identity

Volumes

Volumes Volume Snapshots

Filter ✖ Delete Volume Snapshots

<input type="checkbox"/>	Name	Description	Size	Status	Volume Name	Actions
<input type="checkbox"/>	snapshot for test-inst-snap2	-	10GB	Available	ubuntu-test	<button>Create Volume</button>
<input type="checkbox"/>	test-snap	-	10GB	Available	ubuntu-test	<button>Create Volume</button>
<input type="checkbox"/>	snapshot for test-snaps	-	10GB	Available	ubuntu-test	<button>Create Volume</button>

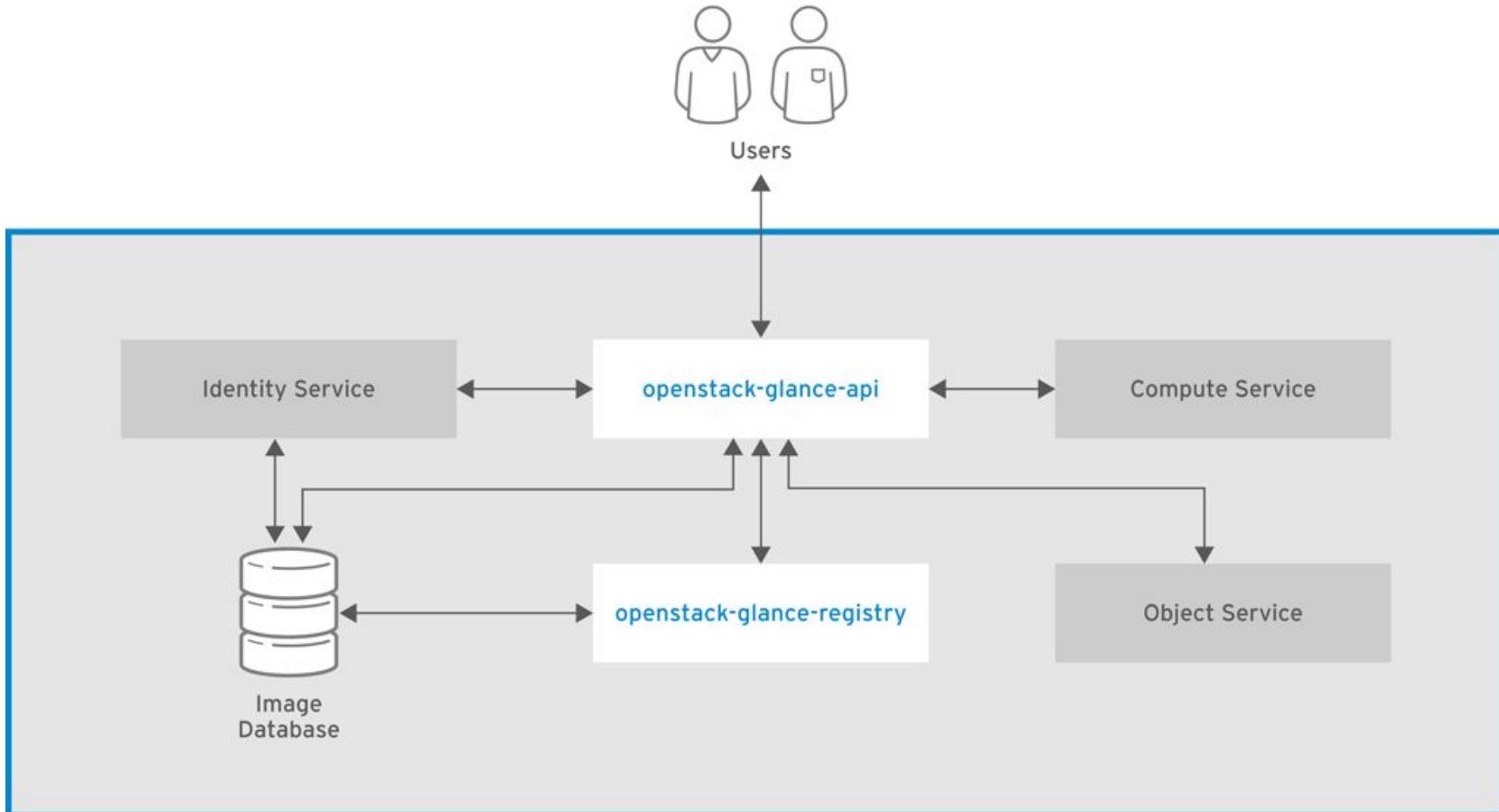
Displaying 3 items

Glance

Сервис предоставления базовых образов ОС.

- Хранит и отдает образы для гипервизоров в Nova.
- В случае использования Docker в Nova отдает заранее построенные образы (т.е Docker-файлы не поддерживаются)

Glance



Glance

Project

Compute

Overview

Instances

Volumes

Images

Access & Security

Network

Orchestration

Data Processing

Object Store

Admin

Identity

Images

Project (5) Shared with Me (0) Public (4) + Create Image × Delete Images

<input type="checkbox"/>	Image Name	Type	Status	Public	Protected	Format	Size	Actions
<input type="checkbox"/>	ubuntu-server-cloud/11.03.2016	Image	Active	Yes	Yes	Raw	247.6 MB	<button>Launch Instance</button>
<input type="checkbox"/>	win-server-2012	Image	Active	No	No	QCOW2	15.6 GB	<button>Launch Instance</button>
<input type="checkbox"/>	ubuntu-trusty	Image	Active	Yes	No	QCOW2	254.7 MB	<button>Launch Instance</button>
<input type="checkbox"/>	cdh5.4.0	Image	Active	Yes	No	QCOW2	3.1 GB	<button>Launch Instance</button>
<input type="checkbox"/>	Cirros	Image	Active	Yes	No	QCOW2	12.7 MB	<button>Launch Instance</button>

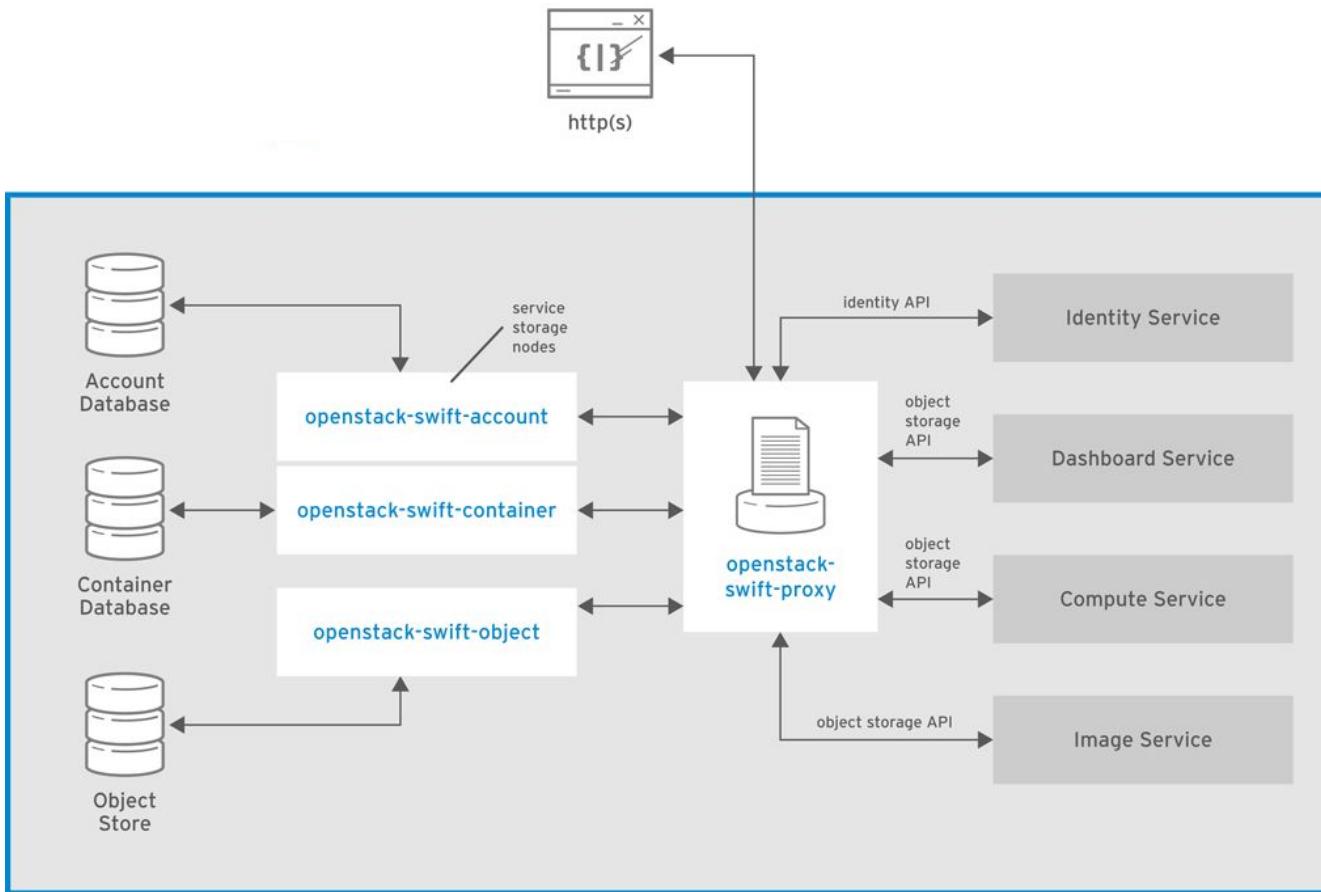
Displaying 5 items

Swift

Объектное хранилище Openstack.

- Позволяет хранить метаданные для файлов (хранятся как xattr).
- Предназначен для использования на обычных машинах.
- Прозрачно масштабируется.
- Поддерживает хранение реплик.
- Может использоваться как замена HDFS в Sahara

Swift



*не указаны:

- openstack-swift-account-auditor
- openstack-swift-container-auditor
- openstack-swift-object-auditor
- openstack-swift-account-replicator
- openstack-swift-container-replicator
- openstack-swift-object-replicator
- openstack-swift-account-updater
- openstack-swift-container-updater
- openstack-swift-object-updater

Swift

Project

Compute

Network

Orchestration

Data Processing

Object Store

Containers

Admin

Identity

Containers

		+ Create Container	Filter	<input type="text"/>	+ Create Pseudo-folder	Upload Object	Delete Objects
container1	Object Count: 1 Size: 37.2 KB Access: Private	View Details	<input type="checkbox"/>	bash_history	37.2 KB	Download	▼

Displaying 1 item

Heat

Сервис оркестрации в Openstack.

- Принимает в себя шаблоны кластеров виртуальных машин (AWS и HOT).
- Предоставляет слой совместимости с метаданными Amazon (неполный).

Heat

Project ▾

Compute

Network

Orchestration ▾

Stacks

Resource Types

Data Processing

Object Store

Admin

Identity

Stacks

<input type="checkbox"/>	Stack Name	Created	Updated	Status	Actions
<input type="checkbox"/>	stress	3 months	Never	Check Complete	<input type="button" value="Check Stack"/> ▾

Displaying 1 item

Heat

Resource Types				
Type	Implementation	Component	Resource	
AWS::AutoScaling::AutoScalingGroup	AWS compatible	AutoScaling	AutoScalingGroup	
AWS::AutoScaling::LaunchConfiguration	AWS compatible	AutoScaling	LaunchConfiguration	
AWS::AutoScaling::ScalingPolicy	AWS compatible	AutoScaling	ScalingPolicy	
AWS::CloudFormation::Stack	AWS compatible	CloudFormation	Stack	
AWS::CloudFormation::WaitCondition	AWS compatible	CloudFormation	WaitCondition	
AWS::CloudFormation::WaitConditionHandle	AWS compatible	CloudFormation	WaitConditionHandle	
AWS::EC2::EIP	AWS compatible	EC2	EIP	
AWS::EC2::EIPAssociation	AWS compatible	EC2	EIPAssociation	
AWS::EC2::Instance	AWS compatible	EC2	Instance	
AWS::EC2::InternetGateway	AWS compatible	EC2	InternetGateway	
AWS::EC2::NetworkInterface	AWS compatible	EC2	NetworkInterface	
AWS::EC2::RouteTable	AWS compatible	EC2	RouteTable	
AWS::EC2::SecurityGroup	AWS compatible	EC2	SecurityGroup	
AWS::EC2::Subnet	AWS compatible	EC2	Subnet	

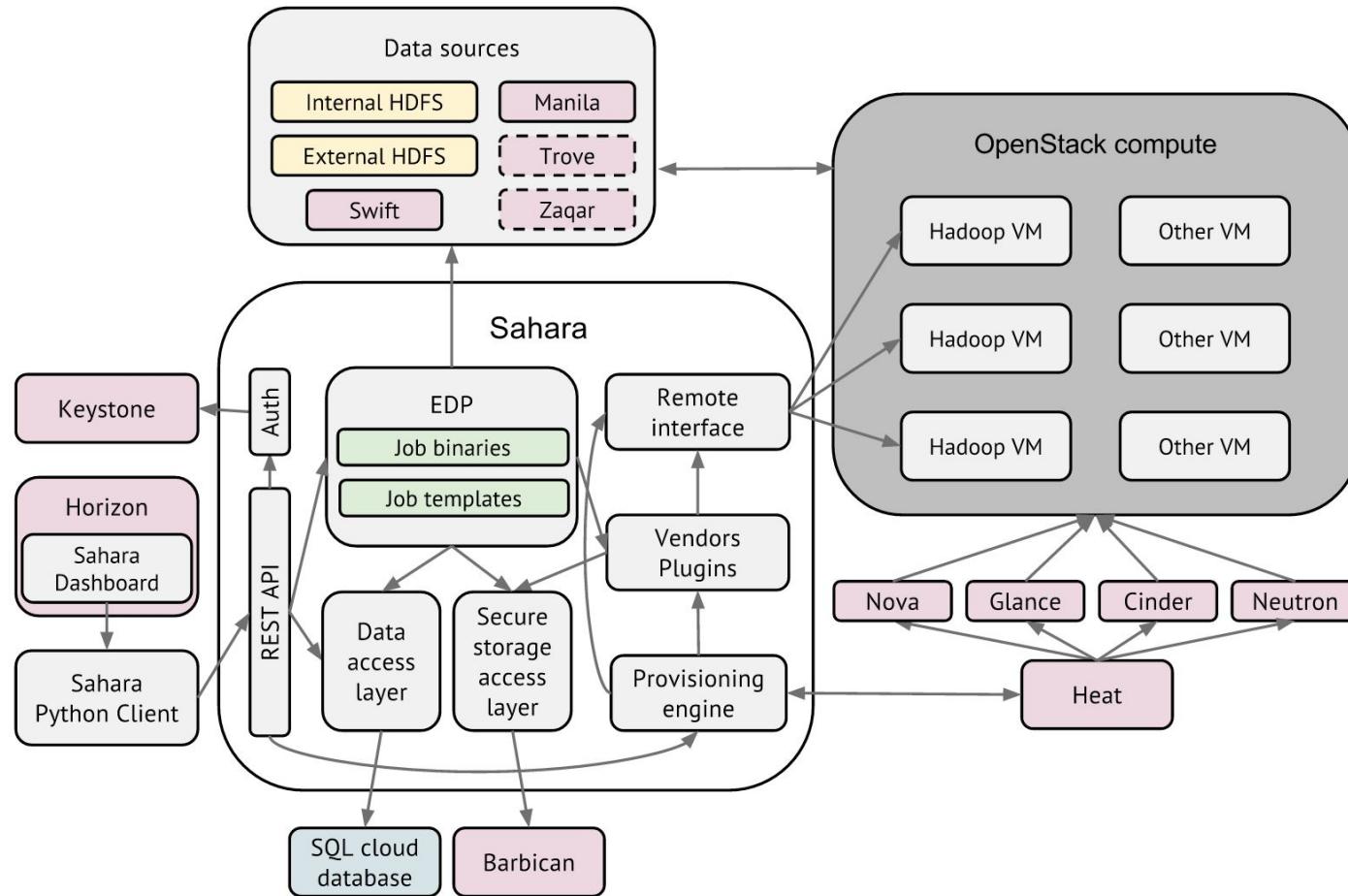
Sahara

Сервис обработки Big Data в PaaS и IaaS ключе.

- Позволяет создавать кластеры с нужными инструментами обработки данных (Hadoop, Spark, Hive, etc)
- Позволяет запускать на исполнение уже написанные задания - не нужно думать про инфраструктуру (закинул-посчитал).

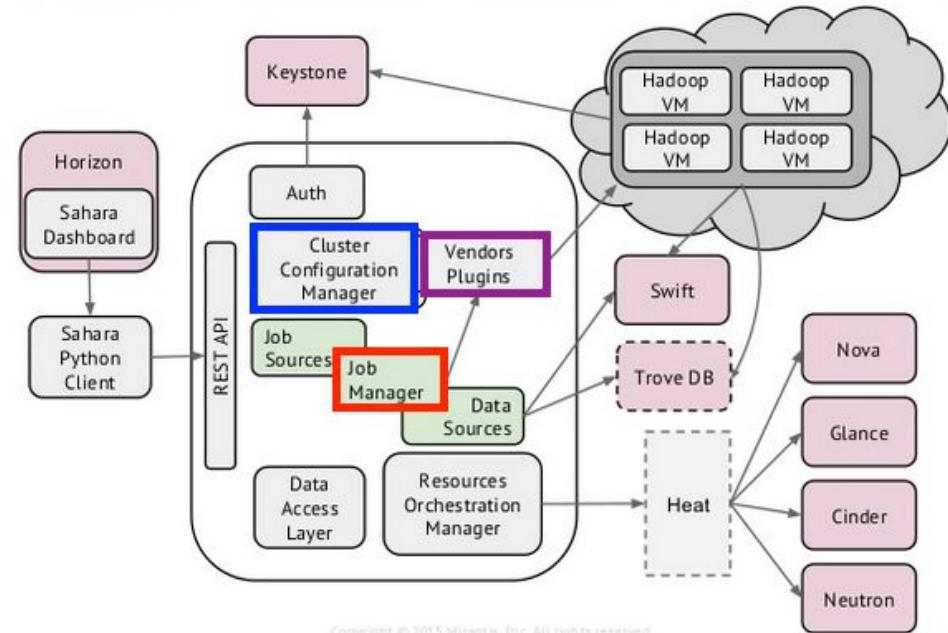
*мы реализовали поддержку Spark-заданий и использование Swift вместо HDFS для Spark

Sahara



Sahara (прошлый год)

- Доработали Cluster Configuration Manager и Vendors plugins для возможности создавать Spark кластеры с возможностью использования Swift.
- Доработали Job Manager для возможности запуска Spark заданий.



Copyright © 2015 Mirantis, Inc. All rights reserved.

Sahara в Horizon

The screenshot shows the OpenStack Horizon dashboard with a dark purple header. The header includes the "openstack" logo, the user "admin", and a dropdown menu.

The main content area has a white background. On the left, there's a sidebar with a "Project" dropdown and several categories: Compute, Network, Object Store, Orchestration, and Data Processing. The "Data Processing" category is currently selected, indicated by a red border around its name and a red bar at the bottom of the sidebar.

The main content area has a title "Data Processing Guides". Below it, there's a text block explaining the process of creating a cluster using Node Group Templates and Cluster Templates, followed by a "Cluster Creation Guide" button.

Further down, there's another text block about running a job, defining input and output paths, and creating a Job Template. This is followed by a "Job Execution Guide" button.

The bottom of the sidebar lists several navigation links: Guides (which is the active tab), Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, and Job Binaries.

Sahara в Horizon

The screenshot shows the OpenStack Horizon dashboard with a dark purple header. The header includes the 'openstack' logo, a user dropdown for 'admin', and another user dropdown for 'admin'. On the left, there's a navigation sidebar with sections for Project, Compute, Network, Object Store, Orchestration, Data Processing (which is currently selected), Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, and Job Templates. The main content area has a title 'Guided Cluster Creation'. It contains three numbered steps:

1. The first step is to determine which type of cluster you want to run. You may have several choices available depending on the configuration of your system. Click on "choose plugin" to bring up the list of data processing plugins. There you will be able to choose the data processing plugin along with the version number. Choosing this up front will allow the rest of the cluster creation steps to focus only on options that are pertinent to your desired cluster type.

[Choose plugin](#)

Current choice: No plugin chosen
2. Next, you need to define the different types of machines in your cluster. This is done by defining a Node Group Template for each type of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines need to be running the "worker" processes. Here, you will define the Node Group Template for your "master" node(s).

[+ Create a Master Node Group Template](#)

Current choice: No Master Node Group Template Created
3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.

[+ Create a Worker Node Group Template](#)

Current choice: No Worker Node Group Template Created

Sahara в Horizon

openstack admin admin

Project Compute Network Object Store Orchestration Data Processing

Clusters Jobs Cluster Templates Node Group Templates Job Templates Job Binaries

Choose plugin and version

Plugin Name * Cloudera Plugin

Select which plugin and version that you want to use to create your cluster.

Cloudera Plugin 5.4.0

Cancel Select

Current choice: No Master Node Group Template Created

3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.

+ Create a Worker Node Group Template

Current choice: No Worker Node Group Template Created

4. Now you need to set the layout of your cluster. By creating a Cluster Template, you will be choosing the number of instances of each

Depending on the you will be able to choose cluster creation steps to

Create a Master Node Group Template

Template for each type processes while another or your "master" node(s).

ИСПРАН

Sahara в Horizon

The screenshot shows the OpenStack Horizon dashboard with a sidebar on the left and a main content area on the right.

Left Sidebar:

- Project
- Compute
- Network
- Object Store
- Orchestration
- Data Processing

Under Data Processing:

- Guides (highlighted)
- Clusters
- Jobs
- Cluster Templates
- Node Group Templates
- Job Templates

Top Navigation:

- openstack
- admin
- admin

Main Content Area:

Guided Cluster Creation

1. The first step is to determine which type of cluster you want to run. You may have several choices available depending on the configuration of your system. Click on "choose plugin" to bring up the list of data processing plugins. There you will be able to choose the data processing plugin along with the version number. Choosing this up front will allow the rest of the cluster creation steps to focus only on options that are pertinent to your desired cluster type.
[Choose plugin](#)
Current choice: Plugin: cdh Version: 5.4.0
2. Next, you need to define the different types of machines in your cluster. This is done by defining a Node Group Template for each type of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines need to be running the "worker" processes. Here, you will define the Node Group Template for your "master" node(s).
[+ Create a Master Node Group Template](#)
Current choice: No Master Node Group Template Created
3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.
[+ Create a Worker Node Group Template](#)
Current choice: No Worker Node Group Template Created

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. On the left, there's a sidebar with navigation links: Project, Compute, Network, Object Store, Orchestration, Data Processing, Cluster Tem..., Node Group Tem..., and Job Tem... At the top, there are two user dropdown menus labeled 'admin'. The main content area is a modal window titled 'Create Node Group Template'. Inside the modal, there's a heading 'Configure Node Group Template *' followed by a grid of parameter categories. The categories are arranged in a grid:

Configure Node Group Template *		Security	CATALOGSERVER Parameters	SENTRY Parameters
STATESTORE Parameters	KMS Parameters	MASTER Parameters	FLUME Parameters	
HUE Parameters	JOBHISTORY Parameters	HDFS_GATEWAY Parameters	IMPALAD Parameters	
REGIONSERVER Parameters	DATANODE Parameters	OOZIE Parameters		
HIVEMETASTORE Parameters	SQOOP Parameters	RESOURCEMANAGER Parameters		
SECONDARYNAMENODE Parameters	NAMENODE Parameters	SOLR Parameters		
KS_INDEXER Parameters	JOURNALNODE Parameters	SPARK_ON_YARN Parameters		
HIVESERVER Parameters	ZOOKEEPER Parameters	YARN_GATEWAY Parameters		
NODEMANAGER Parameters	WEBHCAT Parameters			

Below the grid, there's a 'Template Name *' input field containing a single vertical bar character. To the right of the input field is a note: 'This Node Group Template will be created for: Plugin: cdb'. The background of the main window has some descriptive text and a close button.

Sahara в Horizon

The screenshot shows the OpenStack Horizon interface for creating a Node Group Template. The left sidebar navigation includes Project, Compute, Network, Object Store, Orchestration, Data Processing, Cluster Tem, Node Group Tem, Job Tem, and Job Bi. The main content area is titled "Create Node Group Template". It contains fields for "Template Name" (cloudera-spark54), "Description" (empty), "OpenStack Flavor" (m1.large), "Availability Zone" (nova), "Storage location" (Ephemeral Drive), "Floating IP Pool" (ext-net), and a checkbox for "Proxy Gateway" which is unchecked. Below these fields is a section titled "Processes * ?" with a single checkbox for "IMPALA CATALOG SERVER" which is also unchecked. A large callout box provides detailed information about the Node Group Template creation process, mentioning the plugin (cdh), version (5.4.0), and the fact that it specifies processes to be launched on each instance. It also notes the choice of flavor to determine VM size and storage options like Ephemeral Drive or Cinder Volume.

Template Name *

cloudera-spark54

Description

OpenStack Flavor *

m1.large

Availability Zone ?

nova

Storage location * ?

Ephemeral Drive

Floating IP Pool

ext-net

Proxy Gateway ?

Processes * ?

IMPALA CATALOG SERVER

This Node Group Template will be created for:
Plugin: cdh
Version: 5.4.0

The Node Group Template object specifies the processes that will be launched on each instance. Check one or more processes. When processes are selected, you may set node scoped configurations on corresponding tabs.

You must choose a flavor to determine the size (VCPU, memory and storage) of all launched VMs.

Data Processing provides different storage location options. You may choose Ephemeral Drive or a Cinder Volume to be attached to instances.

depending on the you will be able to choose cluster creation steps to

up Template for each type if processes while another for your "master" node(s).

Template.

number of instances of each

Sahara в Horizon

The screenshot shows the OpenStack Horizon interface for creating a new Sahara cluster. On the left, a sidebar menu lists various services under the 'Data Processing' category. The main panel displays a list of service components with checkboxes. Most checkboxes are checked, indicating they will be included in the cluster. A tooltip provides information about the chosen services.

Selected Services:

- CLOUDERA_MANAGER
- YARN_JOBHISTORY
- IMPALAD
- HBASE_REGIONSERVER
- HBASE_MASTER
- OOZIE_SERVER
- HDFS_DATANODE
- SENTRY_SERVER
- YARN_RESOURCEMANAGER
- HDFS_SECONDARYNAMENODE
- HIVE_METASTORE
- HDFS_NAMENODE
- SOLR_SERVER
- YARN_NODEMANAGER
- KEY_VALUE_STORE_INDEXER
- HDFS_JOURNALNODE
- SPARK_YARN_HISTORY_SERVER
- HIVE_SERVER2
- SQOOP_SERVER
- HIVE_WEBHCAT

Tooltip Content:

depending on the
you will be able to choose
cluster creation steps to

up Template for each type
if processes while another
for your "master" node(s).

Template.

Buttons:

Cancel Create

Number of instances of each

Sahara в Horizon

- Network
- Object Store
- Orchestration
- Data Processing
 - Guides
 - Clusters
 - Jobs
 - Cluster Templates
 - Node Group Templates
 - Job Templates
 - Job Binaries
 - Data Sources
 - Image Registry

configuration of your system. Click on "choose plugin" to bring up the list of data processing plugins. There you will be able to choose the data processing plugin along with the version number. Choosing this up front will allow the rest of the cluster creation steps to focus only on options that are pertinent to your desired cluster type.

[Choose plugin](#)

Current choice: Plugin: cdh Version: 5.4.0

2. Next, you need to define the different types of machines in your cluster. This is done by defining a Node Group Template for each type of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines need to be running the "worker" processes. Here, you will define the Node Group Template for your "master" node(s).

[+ Create a Master Node Group Template](#)

Current choice: Master Node Group Template: cloudera-spark54

3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.

[+ Create a Worker Node Group Template](#)

Current choice: No Worker Node Group Template Created

4. Now you need to set the layout of your cluster. By creating a Cluster Template, you will be choosing the number of instances of each Node Group Template that will appear in your cluster. Additionally, you will have a chance to set any cluster-specific configuration items in the additional tabs on the create Cluster Template form.

[+ Create a Cluster Template](#)

Current choice: No Cluster Template Created

Sahara в Horizon

NODEMANAGER Parameters **WEBHCAT Parameters**

Template Name *
cloudera-spark54slave

Description

This Node Group Template will be created for:
Plugin: cdh
Version: 5.4.0

The Node Group Template object specifies the processes that will be launched on each instance. Check one or more processes. When processes are selected, you may set node scoped configurations on corresponding tabs.

You must choose a flavor to determine the size (VCPU, memory and storage) of all launched VMs.

Data Processing provides different storage location options. You may choose Ephemeral Drive or a Cinder Volume to be attached to instances.

OpenStack Flavor *
m1.large

Availability Zone ?
nova

Storage location * ?
Ephemeral Drive

Floating IP Pool
ext-net

Proxy Gateway ?

you will be able to choose cluster creation steps to

up Template for each type of processes while another for your "master" node(s).

o Template.

number of instances of each specific configuration

Sahara в Horizon

The screenshot shows a modal dialog for creating a new node group template. On the left, a sidebar lists categories: Network, Object Store, Orchestration, Data Processing, Cluster Templates, Node Group Templates, Job Templates, Job Batches, Data Sources, and Image Registry. The 'Data Processing' category is currently selected.

The main content area displays a list of service roles with checkboxes:

- CLOUDERA_MANAGER
- YARN_JOBHISTORY
- IMPALAD
- HBASE_REGIONSERVER
- HBASE_MASTER
- OOZIE_SERVER
- HDFS_DATANODE
- SENTRY_SERVER
- YARN_RESOURCEMANAGER
- HDFS_SECONDARYNAMENODE
- HIVE_METASTORE
- HDFS_NAMENODE
- SOLR_SERVER
- YARN_NODEMANAGER
- KEY_VALUE_STORE_INDEXER
- HDFS_JOURNALNODE
- SPARK_YARN_HISTORY_SERVER
- HIVE_SERVER2
- SQOOP_SERVER
- HIVE_WEBHCAT

On the right side of the dialog, there are several informational text blocks:

- "you will be able to choose cluster creation steps to..."
- "Setup Template for each type of processes while another for your "master" node(s)."
- "Template."
- "Number of instances of each -specific configuration"

At the bottom right of the dialog are two buttons: "Cancel" and "Create".

Sahara в Horizon

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Data Sources

Image Registry

Plugins

Admin

▼

Identity

▼

of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines need to be running the "worker" processes. Here, you will define the Node Group Template for your "master" node(s).

[+ Create a Master Node Group Template](#)

Current choice: Master Node Group Template: cloudera-spark54

3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.

[+ Create a Worker Node Group Template](#)

Current choice: Worker Node Group Template: cloudera-spark54slave

4. Now you need to set the layout of your cluster. By creating a Cluster Template, you will be choosing the number of instances of each Node Group Template that will appear in your cluster. Additionally, you will have a chance to set any cluster-specific configuration items in the additional tabs on the create Cluster Template form.

[+ Create a Cluster Template](#)

Current choice: No Cluster Template Created

5. You are now ready to launch your cluster. When you click on the link below, you will need to give your cluster a name, choose the Cluster Template to use and choose which image to use to build your instances. After you click on "Create", your instances will begin to spawn. Your cluster should be operational in a few minutes.

[+ Launch a Cluster](#)

[Reset Cluster Creation Guide](#)

Sahara в Horizon

2. Now, you need to define the different types of machines in your cluster. This is done by defining a Node Group Template for each type of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines run a "slave" set of processes. You can define multiple Node Group Templates for your "master" node(s).

Create Cluster Template

Details * Node Groups * General Parameters SENTRY Parameters KMS Parameters

FLUME Parameters HUE Parameters YARN Parameters OOZIE Parameters

SQOOP Parameters SOLR Parameters HBASE Parameters HDFS Parameters

KS_INDEXER Parameters SPARK_ON_YARN Parameters ZOOKEEPER Parameters

HIVE Parameters IMPALA Parameters

Template Name *
cl54

This Cluster Template will be created for:
Plugin: cdh
Version: 5.4.0

The Cluster Template object should specify Node Group Templates that will be used to build a Cluster. You can add Node Groups using Node Group Templates on a "Node Groups" tab.

You may set cluster scoped configurations on

Sahara в Horizon

Create Cluster Template

Up Template for each type of processes while another for your "master" node(s).

Cluster Template

Node Group Template

Job Template

Job Binary

Data Source

Image Registry

Admin

Identity

Details * Node Groups * General Parameters SENTRY Parameters KMS Parameters

FLUME Parameters HUE Parameters YARN Parameters OOZIE Parameters

SQOOP Parameters SOLR Parameters HBASE Parameters HDFS Parameters

KS_INDEXER Parameters SPARK_ON_YARN Parameters ZOOKEEPER Parameters

HIVE Parameters IMPALA Parameters

Select a Node Group Template to add:

Select +

Group Name	Template	Count	
cloudera-spark54	cloudera-spark54	1 - +	Remove
cloudera-spark54slave	cloudera-spark54slave	3 - +	Remove

Template for each type of processes while another for your "master" node(s).

Template.

Number of instances of each specific configuration

For a name, choose the your instances will begin

Cancel Create

Sahara в Horizon

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Data Sources

Image Registry

Plugins

Admin

▼

Identity

▼

of machine. A very common case is where you need to have one or more machines running a "master" set of processes while another set of machines need to be running the "worker" processes. Here, you will define the Node Group Template for your "master" node(s).

[+ Create a Master Node Group Template](#)

Current choice: Master Node Group Template: cloudera-spark54

3. Repeat the Node Group Template creation process, but this time you are creating your "worker" Node Group Template.

[+ Create a Worker Node Group Template](#)

Current choice: Worker Node Group Template: cloudera-spark54slave

4. Now you need to set the layout of your cluster. By creating a Cluster Template, you will be choosing the number of instances of each Node Group Template that will appear in your cluster. Additionally, you will have a chance to set any cluster-specific configuration items in the additional tabs on the create Cluster Template form.

[+ Create a Cluster Template](#)

Current choice: Worker Node Group Template: cl54

5. You are now ready to launch your cluster. When you click on the link below, you will need to give your cluster a name, choose the Cluster Template to use and choose which image to use to build your instances. After you click on "Create", your instances will begin to spawn. Your cluster should be operational in a few minutes.

[+ Launch a Cluster](#)

[Reset Cluster Creation Guide](#)

Sahara в Horizon

Project

Compute

Network

Object Store

Orchestration

Data Processing

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Data Sources

Image Registry

Plugins

Admin

Identity

Reset Cluster Creation

Guided Cluster Creation

1. The first step is to define the basic properties of your cluster. This includes the name, description, and cluster template.

2. Next, you need to specify the base image for your cluster instances. You can choose from a list of available images.

3. Repeat the previous steps for each node group template that will appear in your cluster.

4. Now you need to specify the keypair that will be used to access the cluster instances. Additionally, you can choose a plugin and version for the cluster.

5. You are now ready to launch the cluster. Click "Launch" to start the process.

Launch Cluster

Configure Cluster *

Cluster Name *

Description

Cluster Template *

Base Image *

Keypair ⓘ

Neutron Management Network *

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.

of your system. Click on "choose plugin" to bring up the front will allow the rest of the cluster creation steps to

machine. A very common case is where you need to have, you will define the Node Group Template for your

Group Template that will appear in your cluster. n.

Template to use and choose which image to use to build

Cancel

Launch

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. The top navigation bar has 'admin' selected. On the left, a sidebar menu is open under the 'Compute' heading, showing options like Project, Compute, Network, Object Store, Orchestration, Data Processing, Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, and Job Binaries. The 'Clusters' link in the sidebar is highlighted in red. The main content area is titled 'Clusters'. It features a search bar with 'Name' dropdown and 'Filter' buttons, a 'Cluster Creation Guide' button, and 'Launch Cluster' and 'Delete Clusters' buttons. A table lists one cluster: 'democluster' (Plugin: cdh, Version: 5.4.0, Status: Spawning, Instances Count: 0). A 'Delete Cluster' button is also present in the Actions column for this row. Below the table, it says 'Displaying 1 item'.

<input type="checkbox"/>	Name	Plugin	Version	Status	Instances Count	Actions
<input type="checkbox"/>	democluster	cdh	5.4.0	Spawning	0	<button>Delete Cluster</button>

Displaying 1 item

Sahara в Horizon

openstack admin admin

Project

Compute

Network

Object Store

Orchestration

Data Processing

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Jobs

Status Filter Filter Job Guide Delete Jobs

ID	Job Template	Cluster	Status	Actions
7a9a1ad8-4ded-4d3e-bc26-bc60535bc581	spark-wordcount-template-swift	Not available	Succeeded	Delete Job

Displaying 1 item

Sahara в Horizon

The screenshot shows the OpenStack Horizon dashboard with a dark purple header. The header includes the 'openstack' logo, the user 'admin', and a dropdown menu.

The main content area has a white background and features the title 'Guided Job Execution'. Below it, a step-by-step guide is displayed:

1. First, select which type of job that you want to run. This choice will determine which other steps are required

A button labeled '+ Select type' is present. Below it, a message says 'Current type: No type chosen'.

At the bottom of the main content area, there is a red button labeled 'Reset Job Execution Guide'.

On the left side, there is a sidebar with a dark grey header labeled 'Project'. It contains several dropdown menus:

- Compute (selected)
- Network
- Object Store
- Orchestration
- Data Processing

Below the sidebar, under the 'Guides' section, there are several links:

- Clusters
- Jobs
- Cluster Templates
- Node Group Templates
- Job Templates
- Job Binaries

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. On the left, there's a sidebar with navigation links: Project, Compute, Network, Object Store, Orchestration, Data Processing (which is currently selected), Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, and Job Binaries. The main content area has tabs for 'Job Types' and 'Job Binaries'. A modal dialog box titled 'Choose job type' is open, listing several options: Java, Streaming MapReduce, Hive, MapReduce, Spark (which is checked with a checked checkbox), and Pig. At the bottom right of the modal are 'Cancel' and 'Select' buttons. Below the modal, the main page displays a section for creating a job template, showing a note about no job template created, a step 3 instruction for launching the job, and a 'Launch job' button.

openstack admin admin

Project

Compute

Network

Object Store

Orchestration

Data Processing

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

admin

Choose job type

- Java
- Streaming MapReduce
- Hive
- MapReduce
- ✓ Spark
- Pig

Cancel Select

Create a job template

Job template: No job template created

3. Launch your job. When launching, you may need to choose your input and output data sources. This is where you would also add any special configuration values, parameters, or arguments that you need to pass along to your job.

Launch job

Reset Job Execution Guide

Sahara в Horizon

The screenshot shows the 'Guided Job Execution' page within the OpenStack Sahara interface. On the left, a sidebar menu lists categories like Project, Compute, Network, Object Store, Orchestration, and Data Processing. Below these are sections for Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, and Job Templates. The main content area is titled 'Guided Job Execution' and contains three numbered steps:

1. First, select which type of job that you want to run. This choice will determine which other steps are required.
A button labeled '+ Select type' is shown. Below it, the text 'Current type: Spark' is displayed.
2. Define your Job Template. This is where you choose the type of job that you want to run (Pig, Java Action, Spark, etc) and choose or upload the files necessary to run it. The inputs and outputs will be defined later.
A button labeled '+ Create a job template' is shown. Below it, the text 'Job template: No job template created' is displayed.
3. Launch your job. When launching, you may need to choose your input and output data sources. This is where you would also add any special configuration values, parameters, or arguments that you need to pass along to your job.
A button labeled 'Launch job' is shown.

At the bottom of the main content area, there is a red button labeled 'Reset Job Execution Guide'.

Sahara в Horizon

openstack

Project

Compute

Network

Object Store

Orchestration

Data Processing

Cluster Tem

Node Group Tem

Job Tem

Job Bi

Create Job Template

Create Job Template * Libs

Name *

Job Type *

Choose a main binary ?

 +

Description

Create a job template with a specified name.

Select the type of your job:

- Pig
- Hive
- Spark
- MapReduce
- Java Action

Choose or create your main binary. Additional libraries can be added from the "Libs" tab.

For Spark jobs, only a main is required, "libs" are optional.

For MapReduce or Java Action jobs, "mains" are not applicable. You are required to add one or more "libs" for these jobs.

You may also enter an optional description for your job template.

admin

Sahara в Horizon

The screenshot shows the 'Guided Job Execution' page within the OpenStack Horizon dashboard. The left sidebar lists various services: Project, Compute, Network, Object Store, Orchestration, and Data Processing. Under Data Processing, 'Guides' is selected. The main content area has a header 'Guided Job Execution'. It contains three numbered steps:

1. First, select which type of job that you want to run. This choice will determine which other steps are required.
A button '+ Select type' is shown, with 'Current type: Spark' indicated below it.
2. Define your Job Template. This is where you choose the type of job that you want to run (Pig, Java Action, Spark, etc) and choose or upload the files necessary to run it. The inputs and outputs will be defined later.
A button '+ Create a job template' is shown, with 'Job template: sparkwc' indicated below it.
3. Launch your job. When launching, you may need to choose your input and output data sources. This is where you would also add any special configuration values, parameters, or arguments that you need to pass along to your job.
A button 'Launch job' is shown.

At the bottom, a red button labeled 'Reset Job Execution Guide' is visible.

Top navigation bar: openstack admin admin

Left sidebar (Data Processing):

- Guides
- Clusters
- Jobs
- Cluster Templates
- Node Group Templates
- Job Templates

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. The top navigation bar includes the OpenStack logo, user account 'admin', and a search bar. The left sidebar lists projects and services: Project, Compute, Network, Object Store, Orchestration, and Data Processing. The 'Compute' service is currently selected. A modal dialog box titled 'Launch Job' is open in the center. It contains two tabs: 'Job *' (selected) and 'Configure *'. Below the tabs is a 'Cluster *' dropdown menu set to 'democluster'. To the right of the dropdown is a descriptive text block with instructions: 'Launch the given job template on a cluster.', 'Choose the cluster to use for the job.', 'Choose the Input Data Source (n/a for Java jobs).', and 'Choose the Output Data Source (n/a for Java jobs.)'. At the bottom of the dialog are 'Cancel' and 'Launch' buttons. A red button labeled 'Reset Job Execution Guide' is visible at the bottom of the sidebar.

openstack admin

Project

Compute

Network

Object Store

Orchestration

Data Processing

Reset Job Execution Guide

Job *

Configure *

Cluster *

democluster

Launch the given job template on a cluster.
Choose the cluster to use for the job.
Choose the Input Data Source (n/a for Java jobs).
Choose the Output Data Source (n/a for Java jobs).

Cancel Launch

Cluster Temp

Node Group Templates

Job Templates

Job Binaries

admin

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface with the 'admin' user logged in. The left sidebar navigation includes 'Project' (Compute, Network, Object Store, Orchestration, Data Processing), 'Guides' (Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, Job Binaries, Data Sources, Image Registry, Plugins), and 'Admin' (Identity). The main content area is titled 'Guided Job' and displays a configuration form for a SparkWordCount job.

Main Class: com.cloudera.sparkwordcount.SparkWordCount

JavaOpts: (empty field)

Configuration:

Name	Value	Action
fs.swift.service.sahara.username	admin	Remove
fs.swift.service.sahara.password	*****	Remove
edp.spark.adapt_for_swift	True	Remove

Arguments:

Value	Action
swift://vg-cont/input	Remove
2	Remove

Buttons: Add, Cancel, Launch

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. The top navigation bar includes the OpenStack logo, user authentication dropdowns for 'admin', and a search bar. On the left, a sidebar menu is open under the 'Compute' category, showing options like 'Project', 'Network', 'Object Store', 'Orchestration', and 'Data Processing'. Below these, there are links for 'Guides', 'Clusters', and 'Jobs', where 'Jobs' is currently selected and highlighted in red. The main content area is titled 'Jobs' and displays a table of running jobs. The table has columns for 'ID', 'Job Template', 'Cluster', 'Status', and 'Actions'. Two job entries are listed:

ID	Job Template	Cluster	Status	Actions
7a9a1ad8-4ded-4d3e-bc26-bc60535bc581	spark-wordcount-template-swift	Not available	Succeeded	<button>Delete Job</button>
dc9191f6-13b3-4047-949f-c5d27d50c49a	spark-wordcount-template-swift	democluster	RUNNING	<button>Delete Job</button>

At the bottom of the table, a message indicates 'Displaying 2 items'.

Sahara в Horizon

The screenshot shows the OpenStack Horizon web interface. The top navigation bar includes the OpenStack logo, user authentication for 'admin', and a navigation menu with 'Project' selected. A sidebar on the left lists project categories: Compute, Network, Object Store, Orchestration, and Data Processing, with 'Data Processing' currently active. Under 'Data Processing', there are links for 'Guides', 'Clusters', and 'Jobs'. The main content area is titled 'Jobs' and displays a table of running jobs. The table has columns for ID, Job Template, Cluster, Status, and Actions. Two entries are listed:

<input type="checkbox"/>	ID	Job Template	Cluster	Status	Actions
<input type="checkbox"/>	7a9a1ad8-4ded-4d3e-bc26-bc60535bc581	spark-wordcount-template-swift	Not available	Succeeded	<button>Delete Job</button>
<input type="checkbox"/>	dc9191f6-13b3-4047-949f-c5d27d50c49a	spark-wordcount-template-swift	democluster	Succeeded	<button>Delete Job</button>

Below the table, a message indicates 'Displaying 2 items'.

Спасибо за внимание.