

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

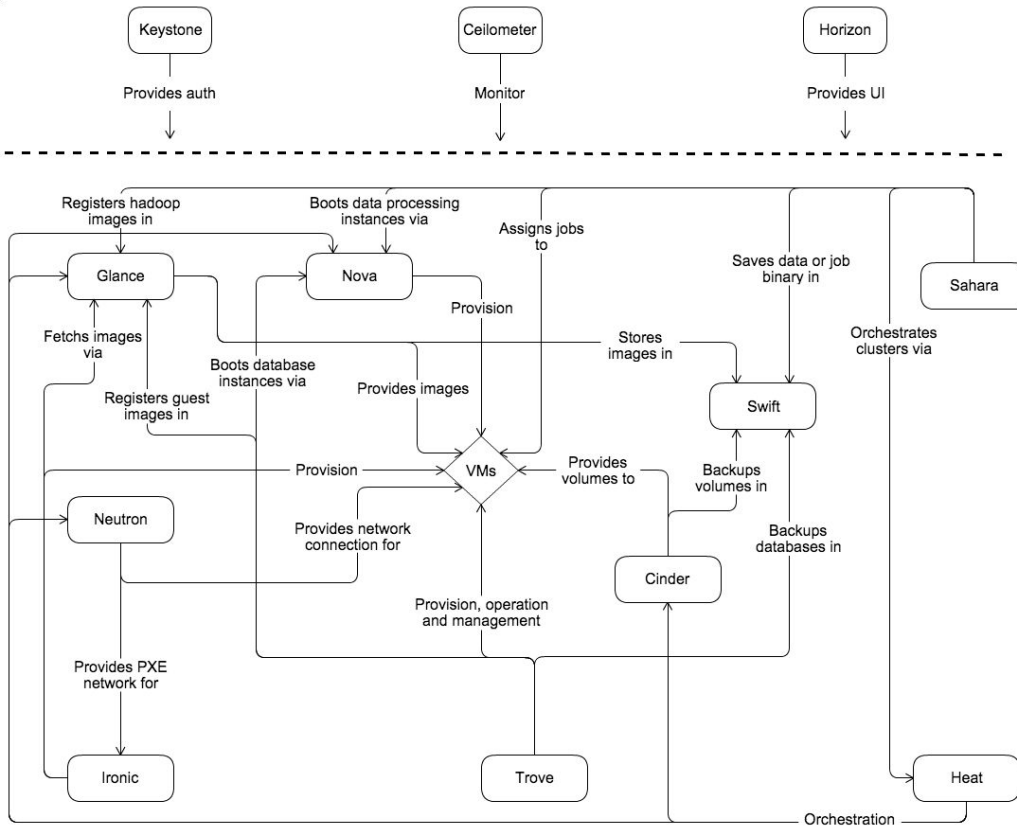
Application Services

-  **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

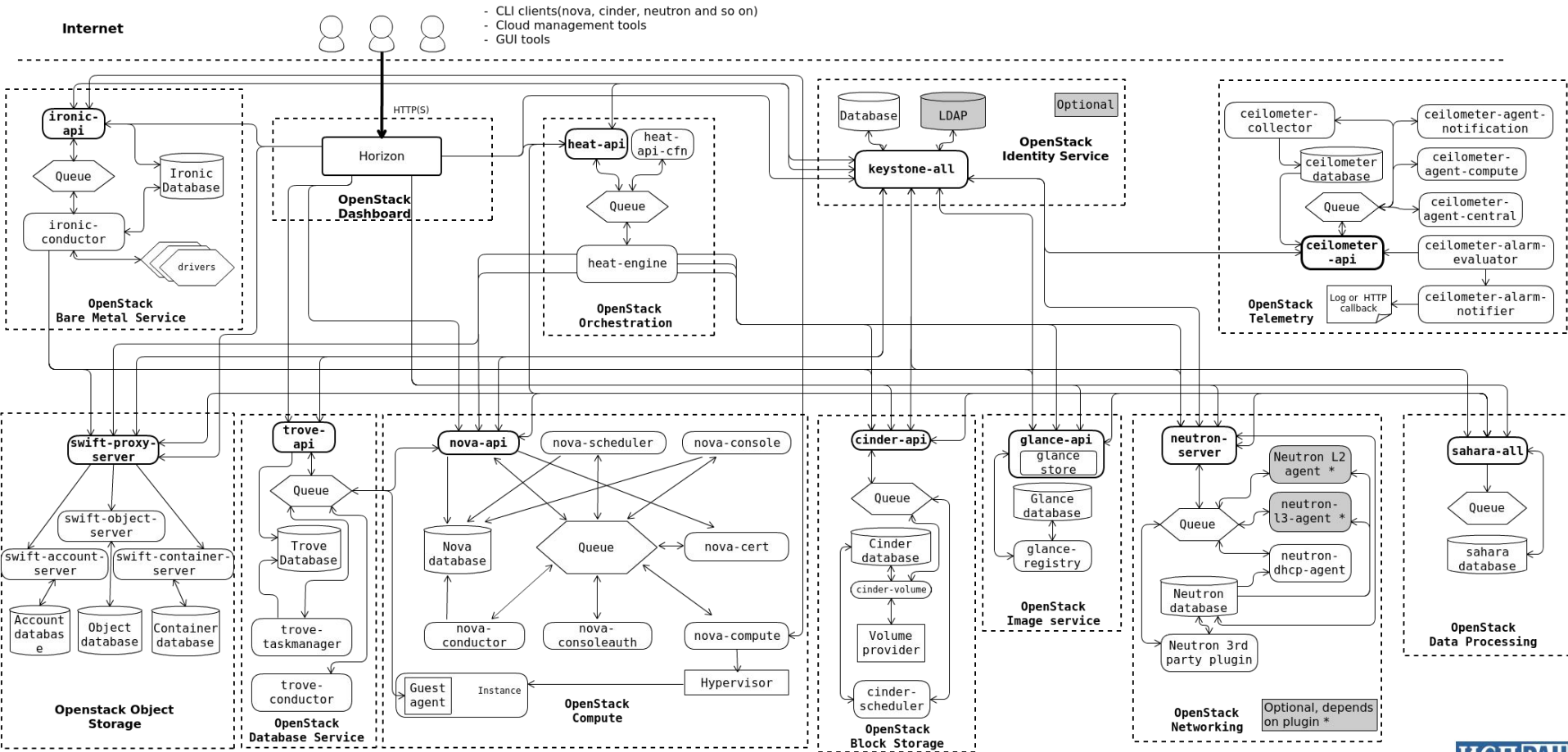


Openstack состоит из т.н. “проектов”.
В реальности их больше, чем на этой
схеме, основные из них:

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

Минимальный необходимый набор —
Keystone, Glance и Nova.

Openstack, полная архитектура

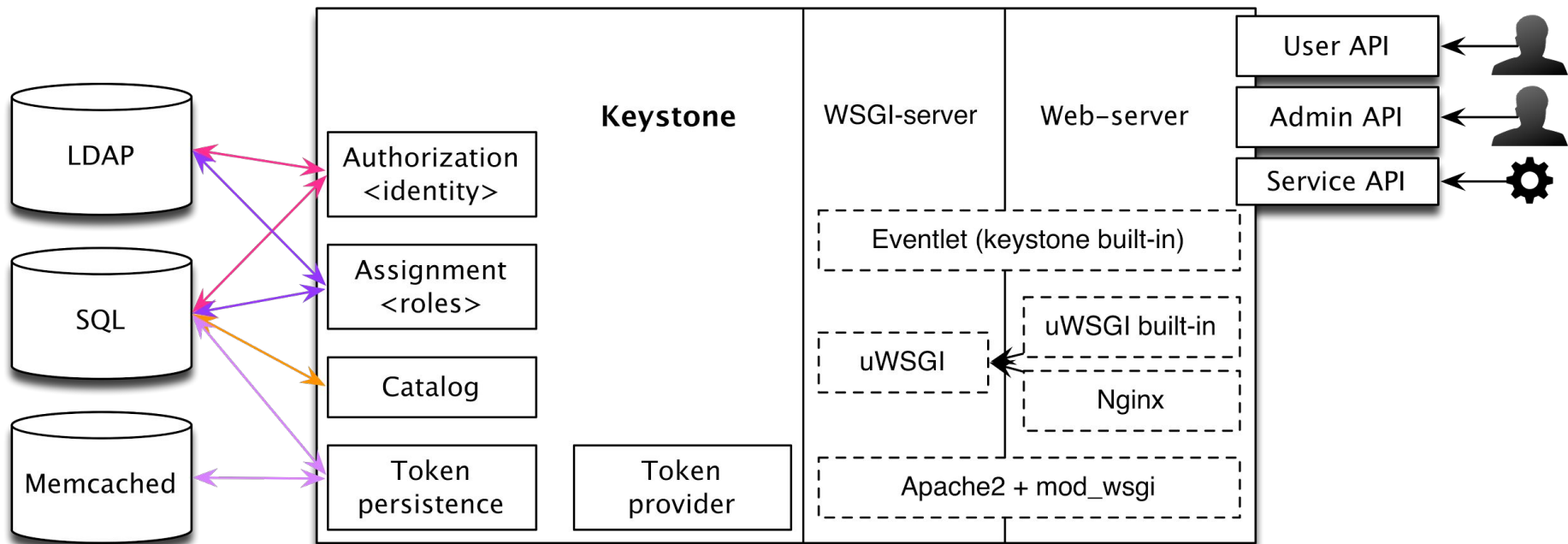


Keystone

Центральный сервис аутентификации и авторизации.

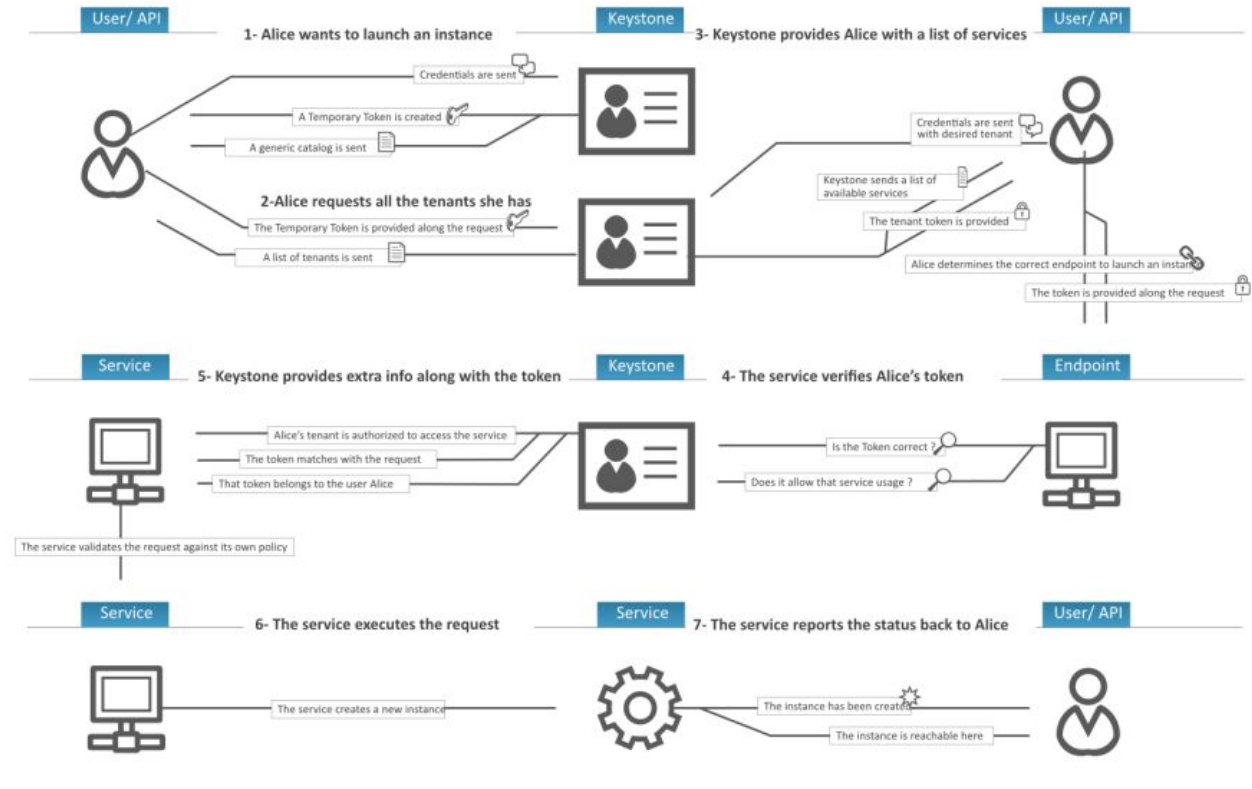
- Обслуживает и пользователей, и сами сервисы
- Работает на основе токенов
- В качестве бэкенда может использовать другие сервисы аутентификации
- Это обычное WSGI-приложение

Keystone



Keystone

The Keystone Identity Manager



Keystone

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	

Keystone

Project

Admin

Identity

Projects

Users

☐

Name

☐

admin

☐

service

☐

demo

☐

computation

☐

outlanders

Displaying 5 items

Edit Project

Project Information *

Project Members

Quota *

All Users

Filter

Q

nova

+

neutron

+

glance

+

cinder

+

sahara

+

heat

+

ceilometer

+

ceph

+

linuxcontest

+

clouddemo

+

Project Members

Filter

Q

admin

member

-

shaman

member

-

al

member

-

fomin

member

-

Q

+ Create Project

x Delete Projects

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 v

Orchestration v

Data Processing v

Object Store v

Admin v

Identity v

Access & Security

[Security Groups](#) [Key Pairs](#) [Floating IPs](#) [API Access](#)

Filter



+ Create Key Pair

⬇ Import Key Pair

✕ Delete Key Pairs

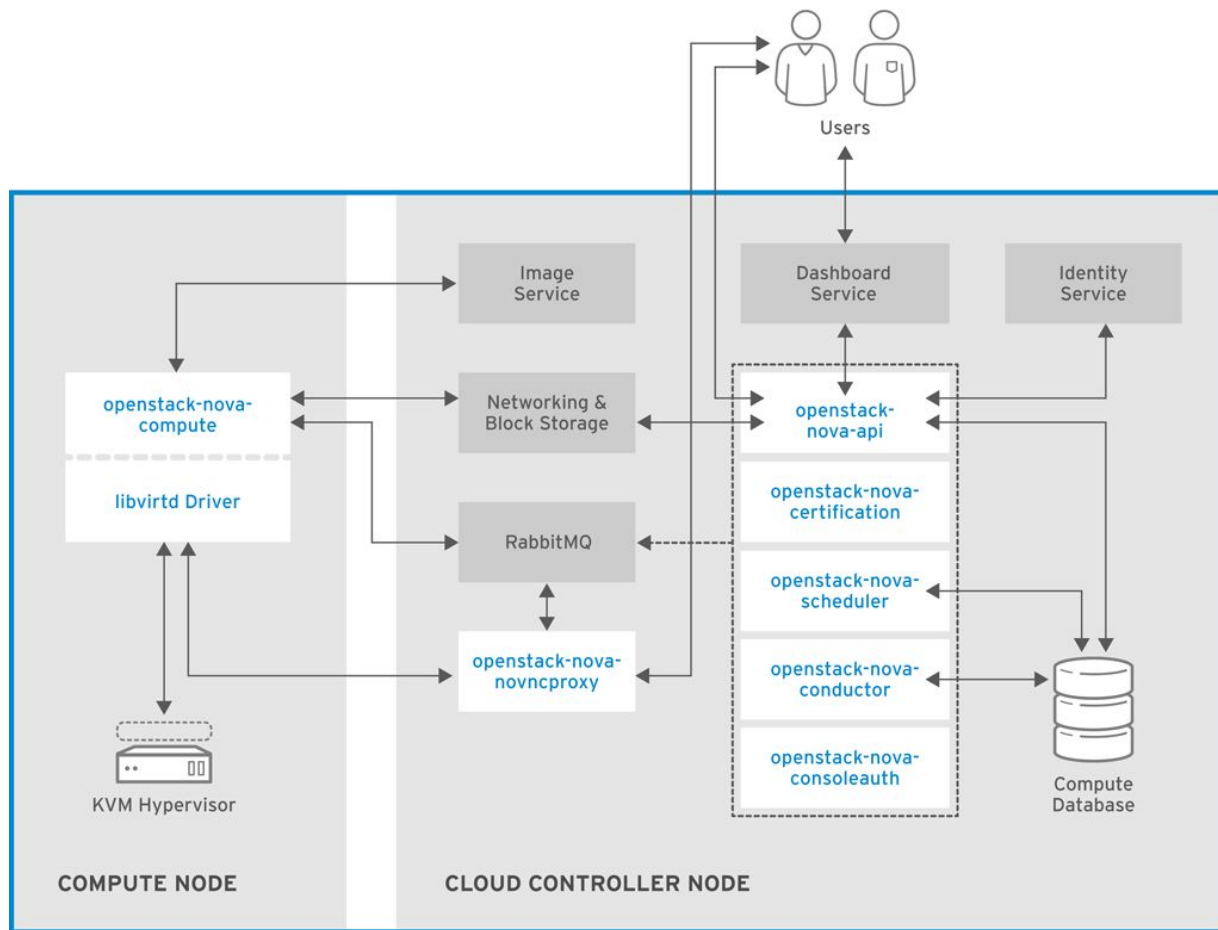
<input type="checkbox"/>	Key Pair Name	Fingerprint	Actions
<input type="checkbox"/>	al	76:c2:ad:0a:1a:bd:3c:fb:f1:a8:bf:6d:5c:ef:f3:13	Delete Key Pair
<input type="checkbox"/>	alexander	dd:88:7b:5b:9e:e8:dd:4e:63:2a:95:be:46:38:c3:63	Delete Key Pair
<input type="checkbox"/>	borisenko	ff:68:b8:02:2c:e4:d2:d9:0f:0e:03:43:34:7f:12:a4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

Launch Instance

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

Availability Zone
nova

Instance Name *

Flavor * ?
func.testing

Instance Count * ?
1

Instance Boot Source * ?
Select source

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

Nova

The screenshot shows the OpenStack Nova web interface. On the left is a sidebar with navigation links: Project, Compute, Overview, Instances, Volumes, Images, Access & Security, Network, Orchestration, Data Processing, Object Store, Admin, and Identity. The main area displays the 'Launch Instance' dialog box, which is a modal window with a title bar and a close button. The dialog has five tabs: Details (selected), Access & Security, Networking, Post-Creation, and Advanced Options. The 'Details' tab contains a 'Key Pair' section with a dropdown menu showing 'al' and a '+' button to add more. Below this is a 'Security Groups' section with a checkbox labeled 'default' that is checked. To the right of these sections is a descriptive text: 'Control access to your instance via key pairs, security groups, and other mechanisms.' At the bottom right of the dialog are 'Cancel' and 'Launch' buttons. In the background, a table of instances is visible, with columns for Instance Name, State, Time since created, and Actions. The table lists two instances: 'testintra' and 'test-resolv'. The 'testintra' instance is in the 'Stopped' state and was created 2 days, 23 hours ago. The 'test-resolv' instance is in the 'Shut Down' state and was created 5 days, 2 hours ago. The 'Actions' column for each instance contains buttons for 'Edit Instance' and 'Start Instance'.

openstack admin

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

Instance Name	State	Time since created	Actions
testintra	Stopped	2 days, 23 hours	Edit Instance
test-resolv	Shut Down	5 days, 2 hours	Start Instance

Nova

Launch Instance

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

Selected networks

- NIC:1 ↔ ispras (e5134af5-8fa5-48a1-9c06-37a803aa68d)
- NIC:2 ↔ test-net (b70a639c-8467-4cd9-afbf-1c12c28ffa51b)

Available networks

- external_network (a2b7af1f-bfc7-4a5c-8dc9-ea0fba7a6a70)

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.

Cancel **Launch**

Background Interface:

Project: Compute

Instances

Instance Name	State	Time since created	Actions
testintra	Running	2 days, 23 hours	Edit Instance
test-resolv	Stopped	5 days, 2 hours	Start Instance

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

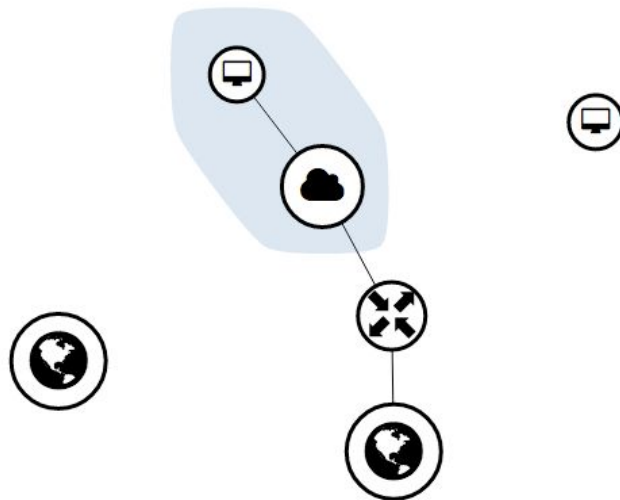
Orchestration

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-cf2c28fba515
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

[+ Create Subnet](#) [x Delete 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	Edit Subnet ^

Displaying 1 item

Ports

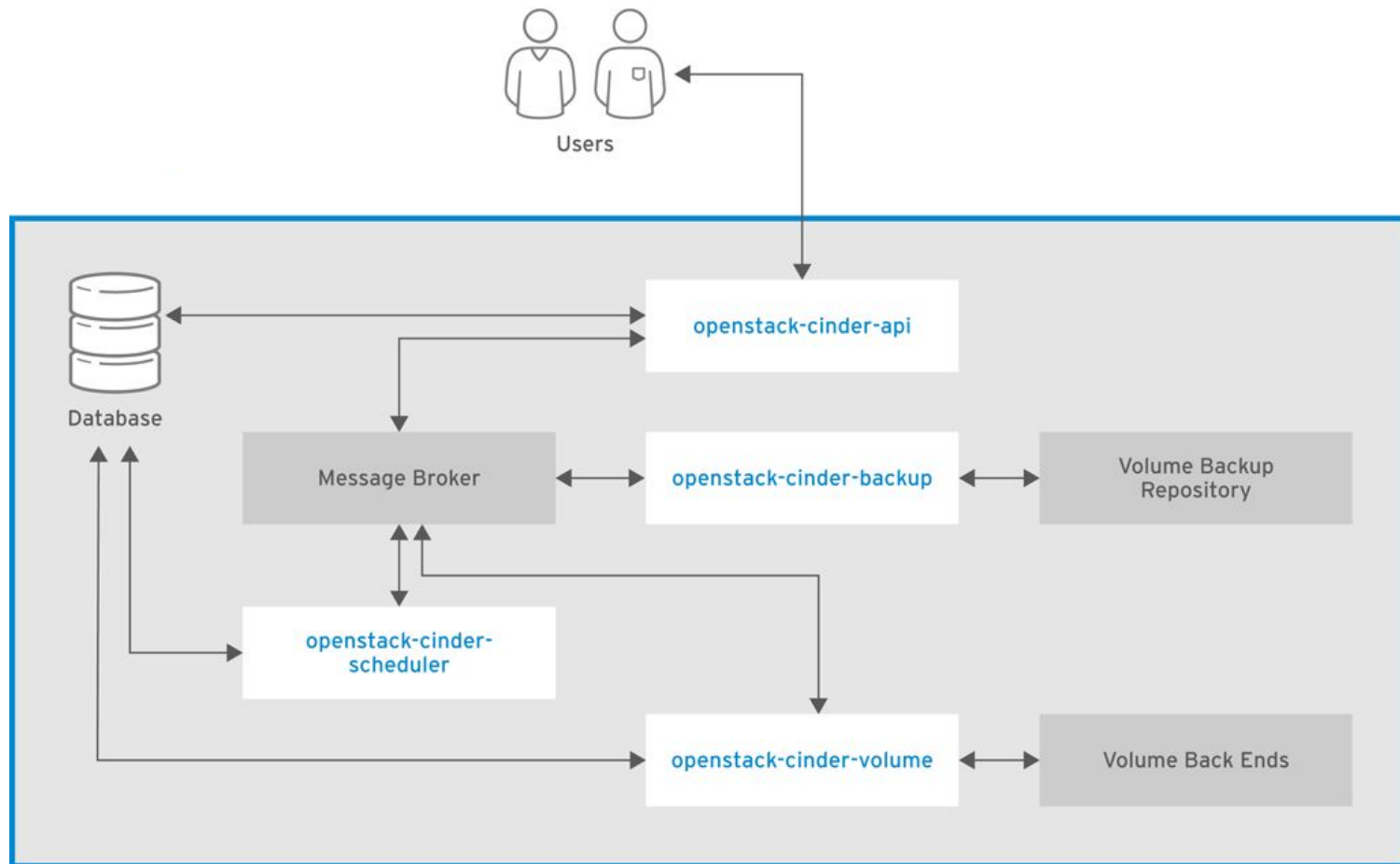
Name	Fixed IPs	Attached Device	Status	Admin State	Actions
(1cc650b7-bc95)	192.168.100.3	compute:nova	Down	UP	Edit Port
(492365de-06ae)	192.168.100.4	compute:nova	Down	UP	Edit Port

Cinder


Сервис предоставления блочных устройств виртуальным машинам.

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

Cinder



Cinder

 admin

admin

Project ^

Compute ^

Overview

Instances

Volumes

Images

Access & Security

Network v

Orchestration v

Data Processing v

Object Store v

Admin v

Identity v

Volumes

Volumes [Volume Snapshots](#)


Filter

<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"/> <input type="button" value="v"/>

Displaying 1 item

ИСПРАИ

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	Create Volume ▾
<input type="checkbox"/>	test-snap	-	10GB	Available	ubuntu-test	Create Volume ▾
<input type="checkbox"/>	snapshot for test-snaps	-	10GB	Available	ubuntu-test	Create Volume ▾

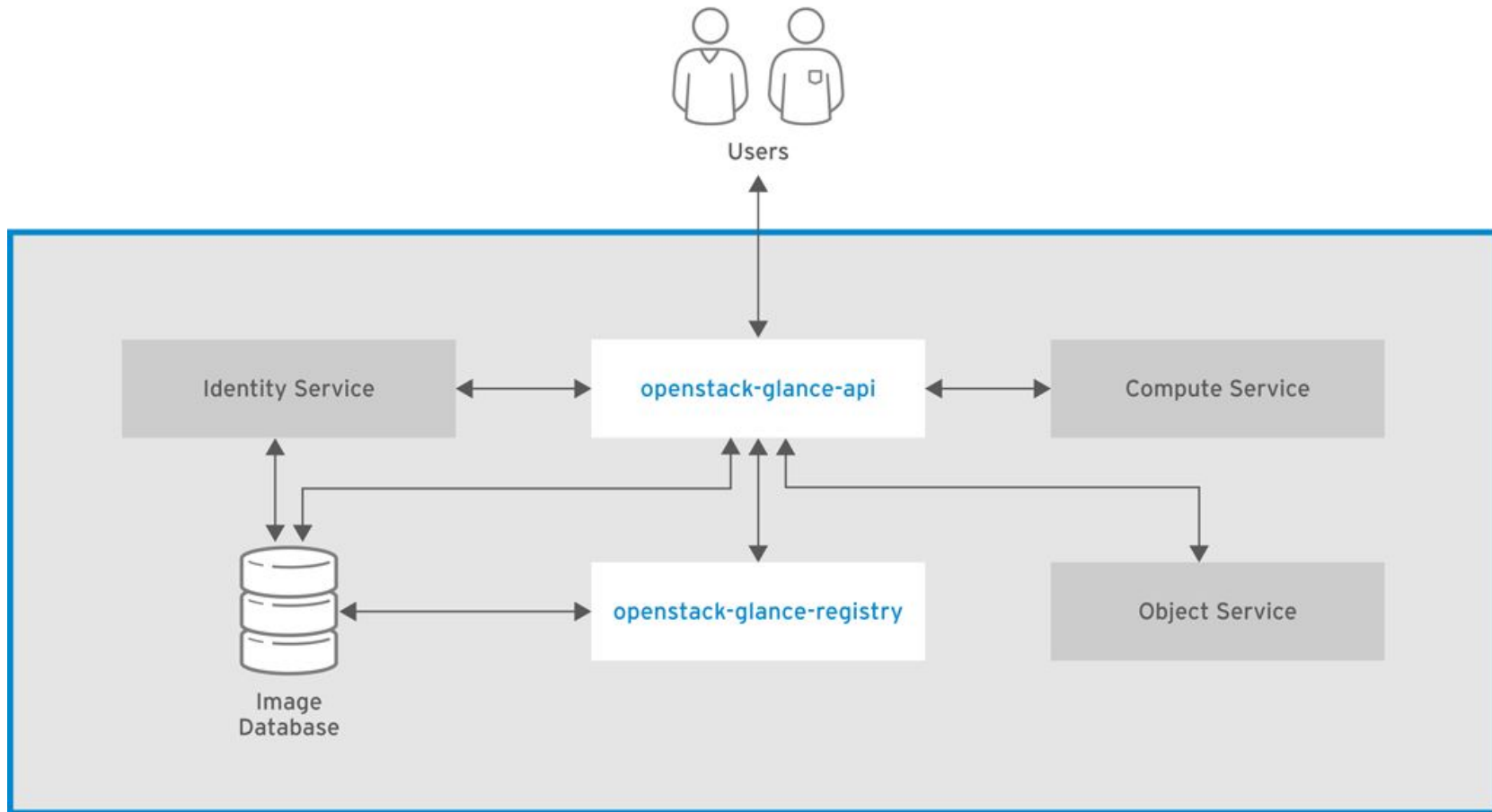
Displaying 3 items

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 v

Orchestration v

Data Processing v

Object Store v

Admin v

Identity v

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	Launch Instance ▼
<input type="checkbox"/>	win-server-2012	Image	Active	No	No	QCOW2	15.6 GB	Launch Instance ▼
<input type="checkbox"/>	ubuntu-trusty	Image	Active	Yes	No	QCOW2	254.7 MB	Launch Instance ▼
<input type="checkbox"/>	cdh5.4.0	Image	Active	Yes	No	QCOW2	3.1 GB	Launch Instance ▼
<input type="checkbox"/>	Cirros	Image	Active	Yes	No	QCOW2	12.7 MB	Launch Instance ▼

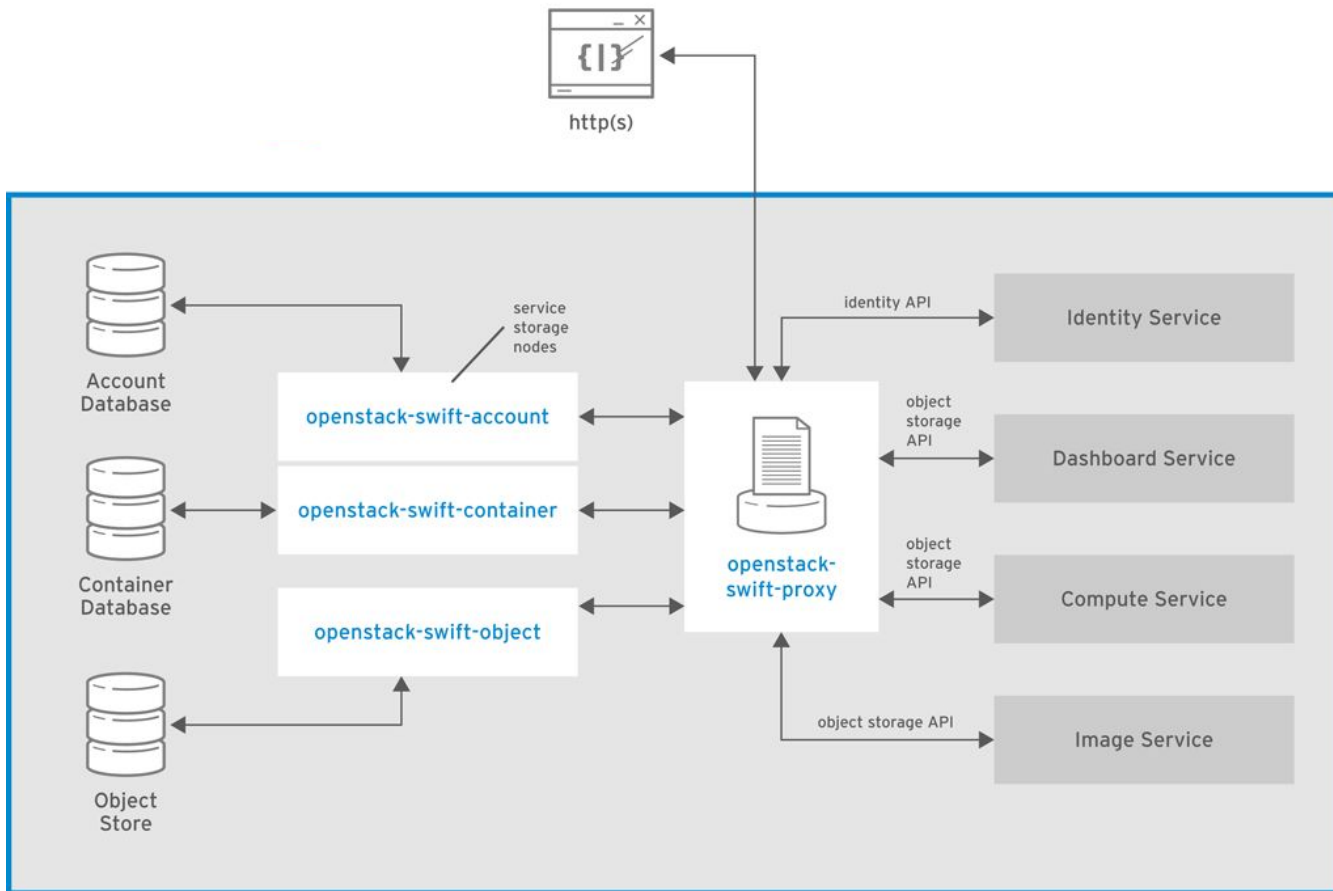
Displaying 5 items

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 v

Network v

Orchestration v

Data Processing v

Object Store ^

Containers

Admin v

Identity v

Containers

+ Create Container

Filter

+ Create Pseudo-folder

Upload Object

✕ Delete Objects

container1	Object Count: 1 Size: 37.2 KB Access: Private	<div>View Details ▾</div>	<input type="checkbox"/>	bash_history	37.2 KB	<div>Download ▾</div>
------------	---	---------------------------	--------------------------	--------------	---------	-----------------------

Displaying 1 item

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 v

Network v

Orchestration ^

Stacks

Resource Types

Data Processing v

Object Store v

Admin v

Identity v

Stacks

[+ Launch Stack](#)[Preview Stack](#)[Check Stacks](#)[Suspend Stacks](#)[Resume Stacks](#)[x Delete Stacks](#)

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

Displaying 1 item

Heat

Project ^

Compute v

Network v

Orchestration ^

Stacks

Resource Types

Data Processing v

Object Store v

Admin v

Identity v

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

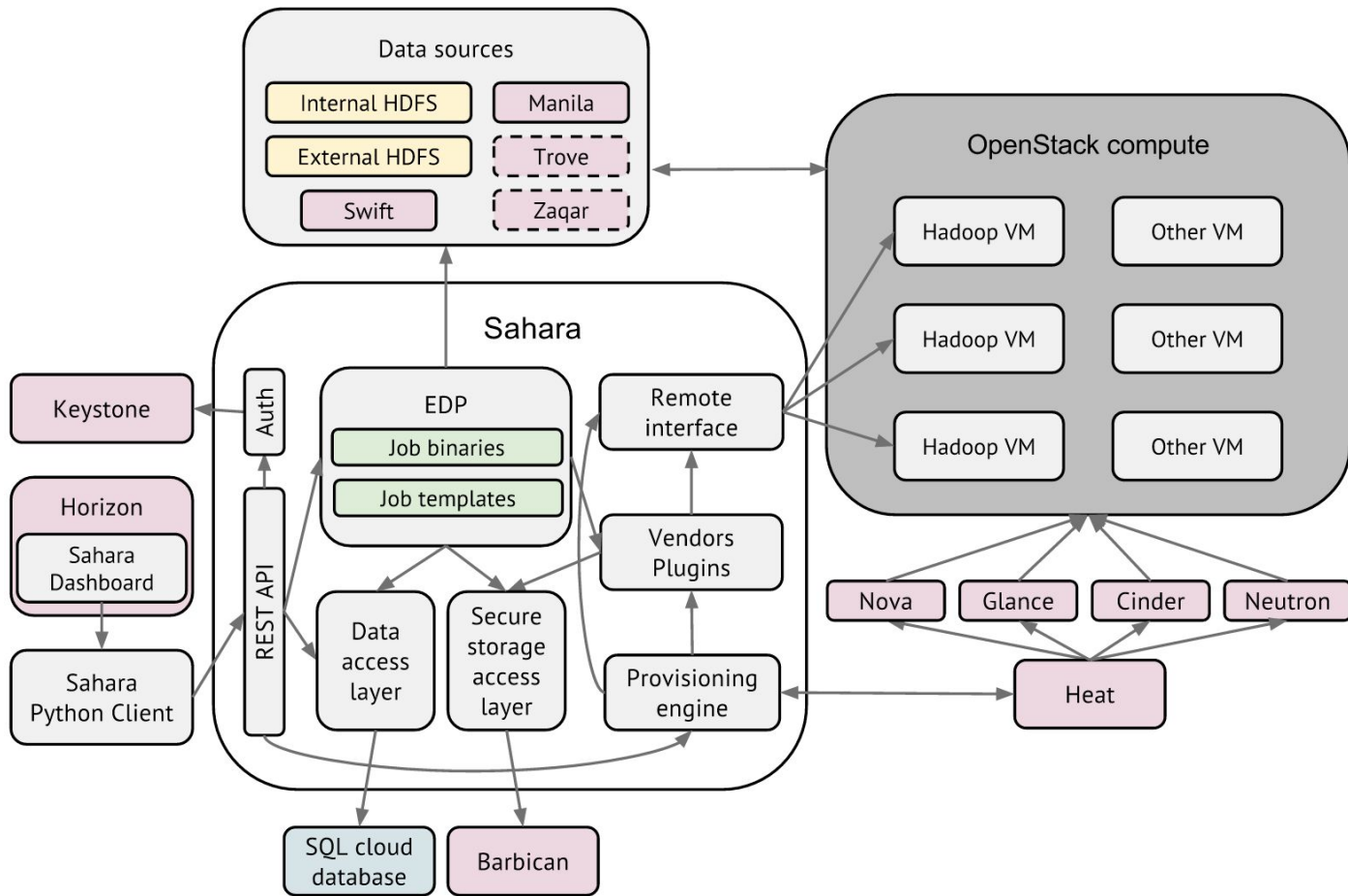
Sahara

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).

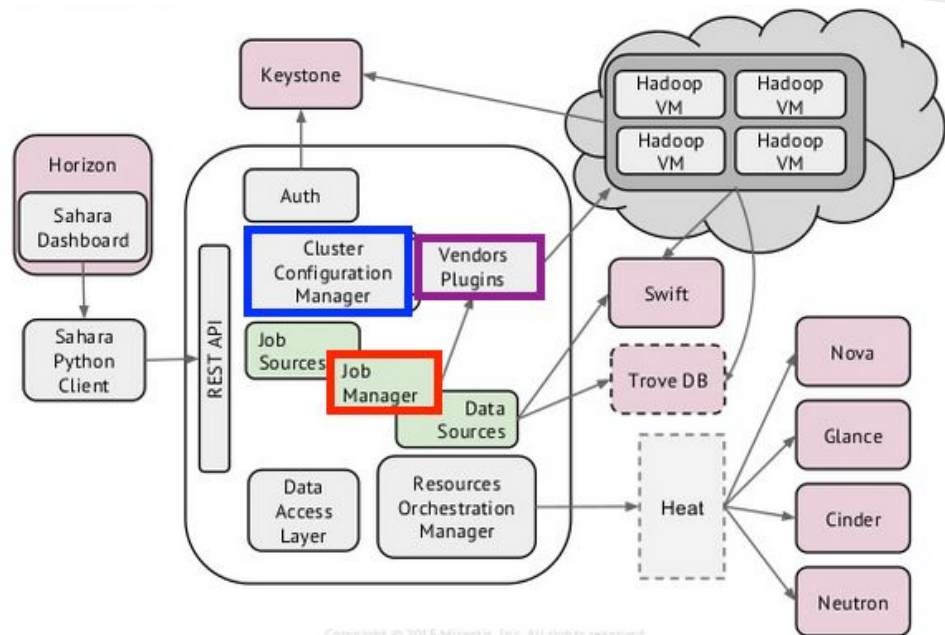
*мы реализовали поддержку Spark-заданий и использование Swift вместо HDFS для Spark

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 web interface. At the top, the OpenStack logo is on the left, and the user 'admin' is logged in on the right. A left-hand navigation menu lists various OpenStack services: Project, Compute, Network, Object Store, Orchestration, and Data Processing. Below this menu, a secondary list of links includes Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, and Job Binaries. The main content area is titled 'Data Processing Guides' and contains a paragraph explaining that each framework requires a cluster of machines. It then provides two links: 'Cluster Creation Guide' and 'Job Execution Guide'.

openstack admin admin

Data Processing Guides

Each of the Data Processing frameworks require a cluster of machines in order to do the work they are assigned. A cluster is formed by creating a set of Node Group Templates, combining those into a Cluster Template and then launching a Cluster. You can do each of those steps manually, or you can follow this guide to help take you through the steps of Cluster creation.

[Cluster Creation Guide](#)

In order to run a Data Processing job, you need to make the files for your program available to the Data Processing system, define where the input and output need to go and create a Job Template that describes how to run your job. Each of those steps can be done manually or you can follow this guide to help take you through the steps to run a job on an existing cluster.

[Job Execution Guide](#)

Guides

Clusters

Jobs

Cluster Templates


Node Group Templates

Job Templates

Job Binaries

ИСП РАН

Sahara в Horizon

 openstack

admin ▾

admin ▾

Project ▾

Compute ▾

Network ▾

Object Store ▾

Orchestration ▾

Data Processing ▾

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

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: 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

The screenshot shows the OpenStack Horizon web interface. On the left is a sidebar with navigation links: Project, Compute, Network, Object Store, Orchestration, and Data Processing. The main content area is titled 'admin' and shows a section for 'Cluster Templates'. A modal dialog box titled 'Choose plugin and version' is open in the center. The dialog has a close button (X) in the top right corner. It contains two dropdown menus: 'Plugin Name *' with 'Cloudera Plugin' selected, and 'Cloudera Plugin' with '5.4.0' selected. A text box to the right of the first dropdown says 'Select which plugin and version that you want to use to create your cluster.' At the bottom right of the dialog are 'Cancel' and 'Select' buttons. In the background, the 'Cluster Templates' section is visible, showing instructions for creating master and worker node group templates. The 'Current choice' for both is 'No Master Node Group Template Created' and 'No Worker Node Group Template Created' respectively. Step 3 of the instructions is highlighted.

openstack admin admin

Choose plugin and version

Plugin Name *

Cloudera Plugin

Cloudera Plugin

5.4.0

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

Cancel Select

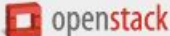
depending on the
you will be able to choose
ster creation steps to

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

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

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

Sahara в Horizon

 admin

Project ^

Compute v

Network v

Object Store v

Orchestration v

Data Processing ^

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Guided Cluster Creation

- 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
- 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).

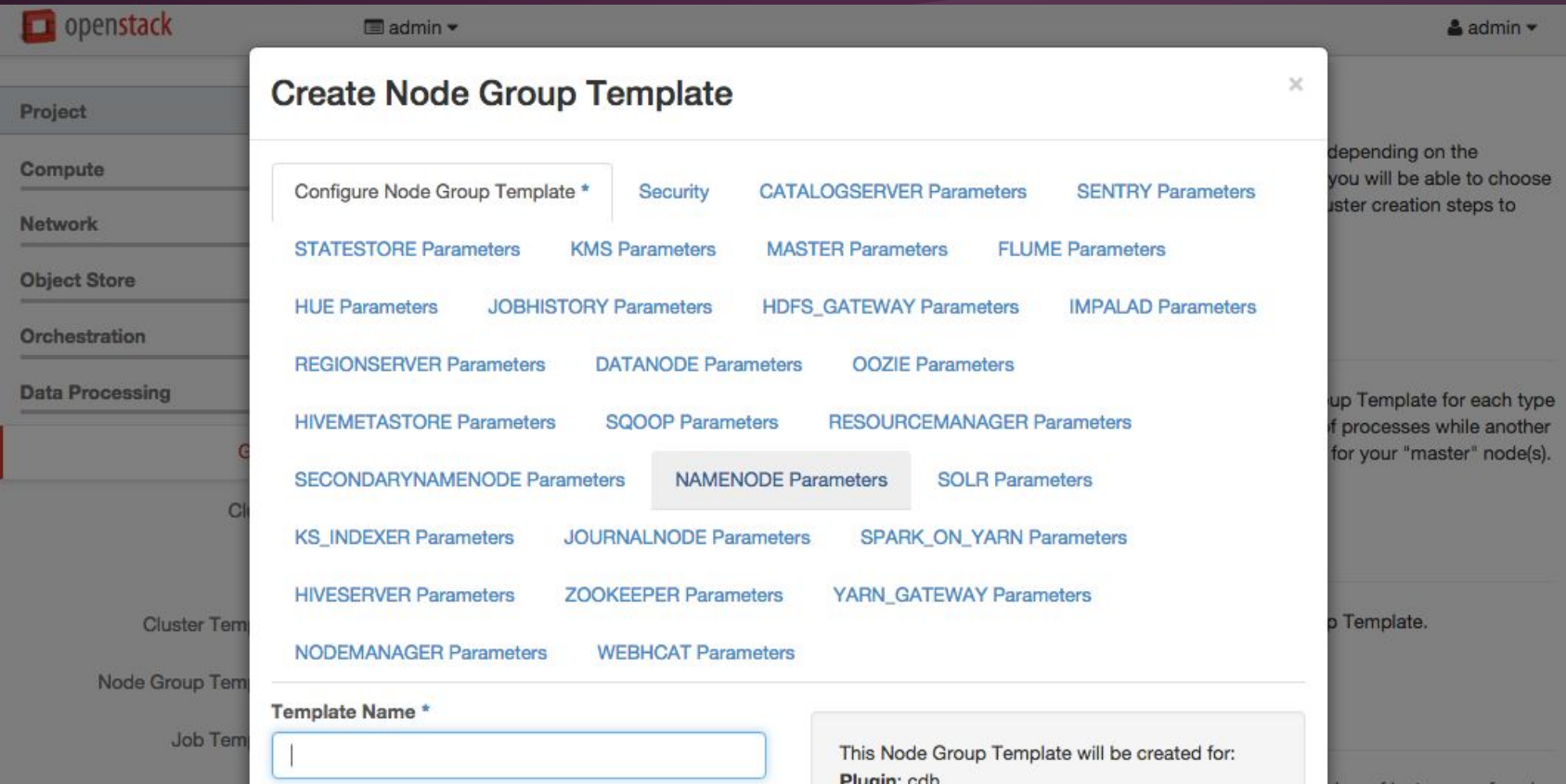
+ Create a Master Node Group Template

Current choice: No Master Node Group Template Created
- 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

Create Node Group Template

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

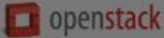
HIVESERVER Parameters ZOOKEEPER Parameters YARN_GATEWAY Parameters

NODEMANAGER Parameters WEBHCAT Parameters

Template Name *

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

Sahara в Horizon



Project

Compute

Network

Object Store

Orchestration

Data Processing

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 CATALOGSERVER

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 (VCPUs, 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.

admin

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

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

o Template.

number of instances of each

ИСП РАН

Sahara в Horizon

The screenshot shows the OpenStack Horizon interface with the 'Data Processing' section selected in the left sidebar. The main content area displays a list of services for the Sahara cluster, each with a checkbox. The following services are checked:

- ☒ 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

At the bottom right, there are two buttons: 'Cancel' and 'Create'.

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

- 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

- 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

- 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

OpenStack Flavor *

m1.large

Availability Zone ?

nova

Storage location * ?

Ephemeral Drive

Floating IP Pool

ext-net

☐ Proxy Gateway ?

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 (VCPUs, 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.

Sahara в Horizon

- ☐ HADOOP_MASTER
- ☐ 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

Cancel

Create

Sahara в Horizon

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Data Sources

Image Registry

Plugins

Admin

Identity

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 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

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

Description

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

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	<div>-</div> <div>+</div> <div>Remove</div>
cloudera-spark54slave	cloudera-spark54slave	3	<div>-</div> <div>+</div> <div>Remove</div>

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

The screenshot displays the Horizon web interface with a sidebar on the left containing navigation links: 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, and Identity. The main content area is titled 'Guided Cluster Creation' and shows a sequence of steps for creating a cluster. A 'Launch Cluster' dialog box is open in the foreground, featuring a 'Configure Cluster' button at the top. The dialog contains several input fields: 'Cluster Name' (set to 'democluster'), 'Description' (empty), 'Cluster Template' (set to 'cl54'), 'Base Image' (set to 'ubuntu_sahara_cloudera_5_4_0'), 'Keypair' (set to 'al_indigo'), and 'Neutron Management Network' (set to 'ext-net'). A 'Launch' button is at the bottom right of the dialog. To the right of the input fields, a text box provides additional information: '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.'

Launch Cluster

Configure Cluster *

Cluster Name *

democluster

Description

Cluster Template *

cl54

Base Image *

ubuntu_sahara_cloudera_5_4_0

Keypair ⓘ

al_indigo

Neutron Management Network *

ext-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

Sahara в Horizon

openstack admin admin

Clusters

Name Filter Filter Cluster Creation Guide + Launch Cluster x Delete Clusters

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

Displaying 1 item

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Sahara в Horizon

openstack admin admin

Jobs

Status Filter Filter Job Guide x Delete Jobs

<input type="checkbox"/>	ID	Job Template	Cluster	Status	Actions
<input type="checkbox"/>	7a9a1ad8-4ded-4d3e-bc26-bc60535bc581	spark-wordcount-template-swift	Not available	Succeeded	Delete Job

Displaying 1 item

Guides

Clusters

Jobs

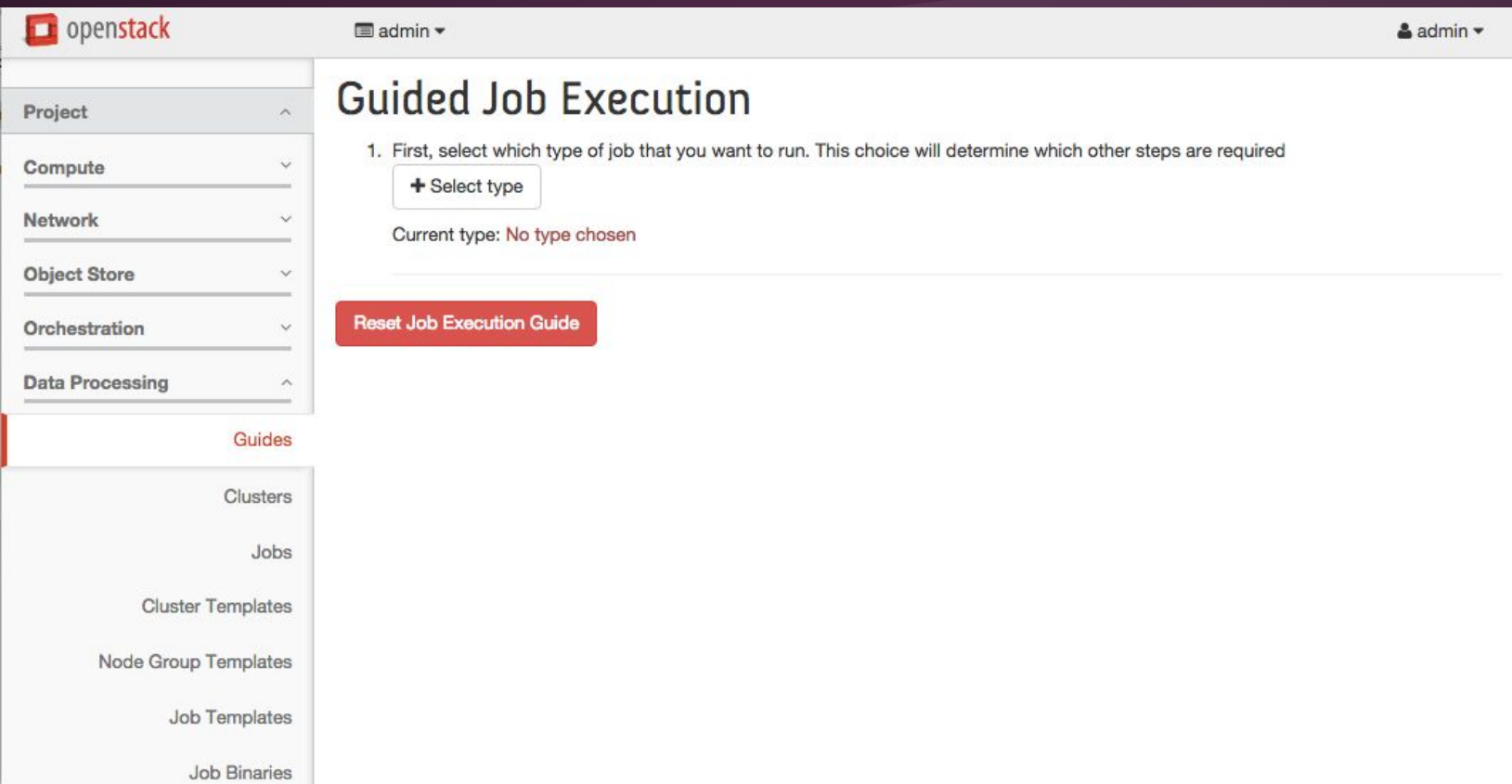
Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Sahara в Horizon



The screenshot shows the OpenStack Horizon web interface. At the top, the OpenStack logo and 'admin' user name are visible. The left sidebar contains a navigation menu with categories like Project, Compute, Network, Object Store, and Data Processing. The main content area is titled 'Guided Job Execution' and contains a step-by-step guide. Step 1 instructs the user to select a job type, with a '+ Select type' button and a 'Current type: No type chosen' status. A red 'Reset Job Execution Guide' button is located below the instructions.

openstack admin admin

Guided Job Execution

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

+ Select type

Current type: No type chosen

Reset Job Execution Guide

Guides

- 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 is a sidebar with navigation links: Project, Compute, Network, Object Store, Orchestration, and Data Processing. The 'Data Processing' section is expanded, showing links to Guides, Clusters, Jobs, Cluster Templates, Node Group Templates, Job Templates, and Job Binaries. The main content area is titled 'Sahara' and shows a 'Create a job template' button. Below it, it says 'Job template: No job template created'. A 'Launch job' button is visible. A 'Reset Job Execution Guide' button is at the bottom. A 'Choose job type' dialog box is open in the center, listing job types: Java, Streaming MapReduce, Hive, MapReduce, Spark (selected with a checkmark), and Pig. The dialog has 'Cancel' and 'Select' buttons. The top of the interface shows the OpenStack logo and the user 'admin'.

openstack admin

Choose job type

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

Cancel Select

Job template: No job template created

Launch job

Reset Job Execution Guide

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.

Sahara в Horizon

openstack admin admin

Guided Job Execution

1. First, select which type of job that you want to run. This choice will determine which other steps are required
[+ Select type](#)
Current type: Spark
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 output will be defined later.
[+ 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](#)

Guides

Clusters

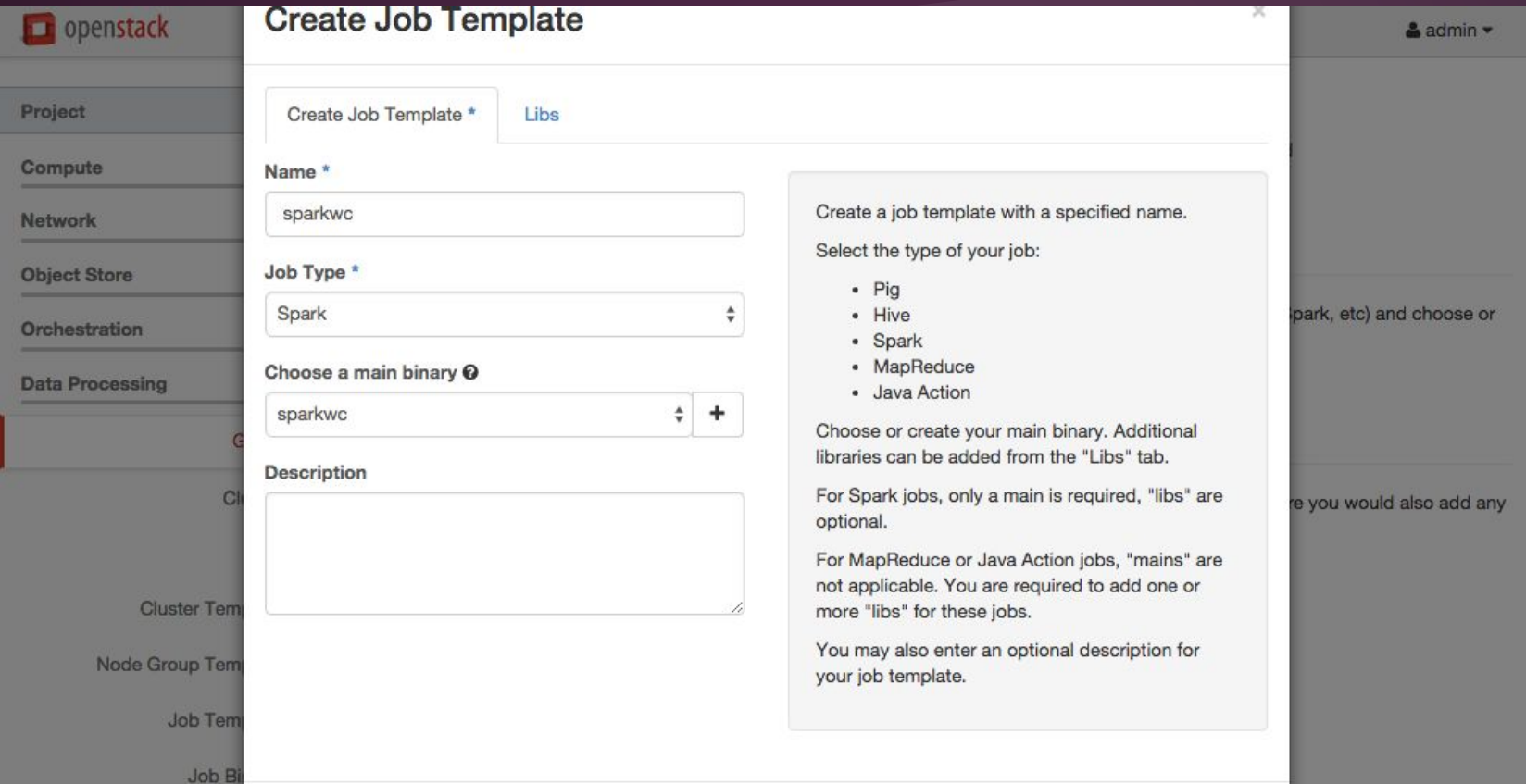
Jobs

Cluster Templates

Node Group Templates

Job Templates

Sahara в Horizon



The screenshot shows the 'Create Job Template' form in the OpenStack Horizon interface. The left sidebar contains navigation links for Project, Compute, Network, Object Store, Orchestration, and Data Processing. The top right shows the user 'admin'. The form has two tabs: 'Create Job Template' (active) and 'Libs'. The 'Create Job Template' tab contains the following fields:

- Name ***: A text input field containing 'sparkwc'.
- Job Type ***: A dropdown menu with 'Spark' selected.
- Choose a main binary ?**: A dropdown menu with 'sparkwc' selected, accompanied by a '+' button to add more binaries.
- Description**: A large text area for entering a description.

To the right of the form is a help box with the following text:

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.

Sahara в Horizon

The screenshot displays the OpenStack Horizon dashboard. The top navigation bar includes the OpenStack logo, a user menu for 'admin', and a role indicator for 'admin'. The left sidebar contains a navigation menu with categories like Project, Compute, Network, Object Store, Orchestration, and Data Processing, along with a 'Guides' section. The main content area is titled 'Guided Job Execution' and presents a three-step process for running a job. Step 1 involves selecting a job type, with 'Spark' currently selected. Step 2 involves defining a job template, with 'sparkwc' currently selected. Step 3 involves launching the job. A red button at the bottom of the steps allows the user to reset the execution guide.

openstack admin admin

Guided Job Execution

1. First, select which type of job that you want to run. This choice will determine which other steps are required.
[+ Select type](#)
Current type: `Spark`
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.
[+ Create a job template](#)
Job template: `sparkwc`
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](#)

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Sahara в Horizon

The screenshot shows the OpenStack Horizon interface with a 'Launch Job' modal dialog box open. The background interface includes the OpenStack logo, a user menu for 'admin', and a sidebar with navigation links: Project, Compute, Network, Object Store, Orchestration, and Data Processing. Below these are links for Cluster Templates, Node Group Templates, Job Templates, and Job Binaries. A 'Reset Job Execution Guide' button is also visible. The 'Launch Job' dialog has a title bar with a close button. It contains two tabs: 'Job *' (selected) and 'Configure *'. Under the 'Job *' tab, there is a 'Cluster *' dropdown menu currently showing 'democluster'. To the right of the dropdown is a text box containing 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). Choose the Output Data Source (n/a for Java jobs).' At the bottom right of the dialog are 'Cancel' and 'Launch' buttons.

openstack admin admin

Project

Compute

Network

Object Store

Orchestration

Data Processing

Cluster Templates

Node Group Templates

Job Templates

Job Binaries

Reset Job Execution Guide

Launch Job

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

Sahara в Horizon

openstack admin

Guided Job

1. First, select a job template to launch. You can also create a new job template.

2. Define your job configuration. You can define job configuration values, parameters, or arguments that will be used to launch the job.

3. Launch your job. You can launch the job immediately or schedule it to run at a later time.

Job template: **com.cloudera.sparkwordcount.SparkWordCount**

Current type: **Spark**

Reset Job Execution

Main Class

com.cloudera.sparkwordcount.SparkWordCount

Java Opts

Configuration

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

Add


Arguments

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

Add

Cancel Launch

Sahara в Horizon

 openstack

admin ▾

admin ▾

Project ▾

Compute ▾

Network ▾

Object Store ▾

Orchestration ▾

Data Processing ▾

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Jobs


Status ▾

Filter

Filter


Job Guide

✕ Delete Jobs

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

Displaying 2 items

Sahara в Horizon

 openstack

admin ▼

admin ▼

Project ^

Compute ▼

Network ▼

Object Store ▼

Orchestration ▼

Data Processing ^

Guides

Clusters

Jobs

Cluster Templates

Node Group Templates

Job Templates

Jobs

Status ▼

Filter

Filter

Job Guide

✕ Delete Jobs

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

Displaying 2 items

Спасибо за внимание.