

Guaranteed Minimum Bandwidth - Feature Demo



Bence Romsics (rubasov)
Balazs Gibizer (gibi)
Miguel Lavalle (mlavalle)



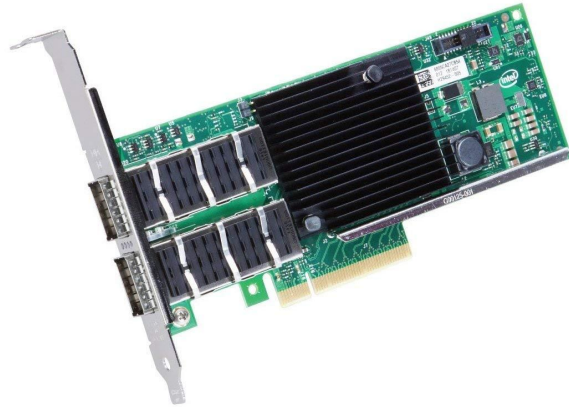
Today's agenda

- Overview
- How it works
- Demo
- Implementation status
- Q&A



Why do you need this?

Network heavy application depends on a minimum amount of bandwidth available for a server running the application



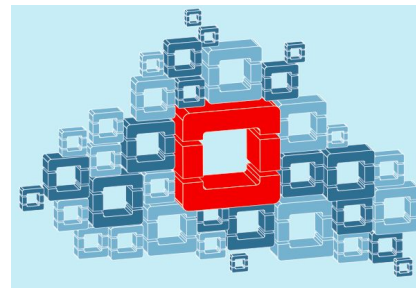
Why we need Placement?

Network bandwidth is a resource managed by Neutron

Placement makes it possible to schedule servers based on **resources** that are **not owned by Nova**.

Placement maintains the **quantitative** and **qualitative resource view** of an OpenStack managed cluster

Placement supports **atomic** resource allocation handling



Compute Resource Provider

name	ubuntu
uuid	d9cf093b-e059-4a6a-91a7-82bddcb39c32

Resource inventories

Quantitative

inventories	DISK_GB	allocation_ratio	1.0
		reserved	0
		total	19
	MEMORY_MB	allocation_ratio	1.5
		reserved	512
		total	7983
	VCPU	allocation_ratio	16.0
		reserved	0
		total	2
name	ubuntu		
uuid	d9cf093b-e059-4a6a-91a7-82bddcb39c32		

Traits

Quantitative

inventories	DISK_GB	allocation_ratio	1.0
		reserved	0
		total	19
	MEMORY_MB	allocation_ratio	1.5
		reserved	512
		total	7983
	VCPU	allocation_ratio	16.0
		reserved	0
		total	2
name	ubuntu		
traits	STORAGE_DISK_SSD HW_CPU_X86_SSE2		
uuid	d9cf093b-e059-4a6a-91a7-82bddcb39c32		

Qualitative

Consumers

```
$ openstack server create \  
  --image cirros-0.3.5-x86_64-disk \  
  --flavor d1 my-vm2
```

```
$ openstack server create \  
  --image cirros-0.3.5-x86_64-disk \  
  --flavor c1 my-vm3
```

inventories	DISK_GB	allocation_ratio	1.0
		reserved	0
		total	19
		used	5
	MEMORY_MB	allocation_ratio	1.5
		reserved	512
		total	7983
		used	768
	VCPU	allocation_ratio	16.0
		reserved	0
		total	2
		used	2
name	ubuntu		
traits	STORAGE_DISK_SSD		
uuid	d9cf093b-e059-4a6a-91a7-82bddcb39c32		

consumes

MEMORY_MB	256
VCPU	1

consumes

DISK_GB	5
MEMORY_MB	512
VCPU	1

consumer_uuid	d00372d4-250b-46fe-af46-fbcb0e9c756
project_id	0de6b35ee2f346f6a33462c370cebc00
user_id	366a19a4ee3042f7a9b113e8ff41b44d

consumer_uuid	0906c1f9-71b2-4469-8843-95882dbc2293
project_id	0de6b35ee2f346f6a33462c370cebc00
user_id	366a19a4ee3042f7a9b113e8ff41b44d

Nested resource providers

aggregates	[]		
inventories	DISK_GB	allocation_ratio	1.0
		reserved	0
		total	19
		used	5
	MEMORY_MB	allocation_ratio	1.5
		reserved	512
		total	7983
		used	768
	VCPU	allocation_ratio	16.0
		reserved	0
		total	2
		used	2
name	ubuntu		
traits	HW_CPU_X86_SSE2 HW_CPU_X86_SSE HW_CPU_X86_MMX HW_CPU_X86_SVM		
uuid	d9cf093b-e059-4a6a-91a7-82bddcb39c32		

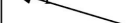
aggregates	[]		
inventories	CUSTOM_MAGIC	allocation_ratio	1.0
		reserved	0
		total	10
		used	0
name	child1		
traits	[]		
uuid	4c191ed8-65ed-4b1f-bdcb-3e32ef2e1d1c		

parent



aggregates	[]		
inventories	CUSTOM_MAGIC	allocation_ratio	1.0
		reserved	0
		total	10
		used	0
name	child2		
traits	[]		
uuid	75d04e60-5056-4a3d-af7e-2e63760f1803		

parent



Modeling the bandwidth resource

inventories	NET_BW_EGRESS_KBPS	total	40000000
	NET_BW_INGRESS_KBPS	total	40000000
name	demo0:NIC Switch agent:ens5		
traits	CUSTOM_PHYSNET_PHYSNET0 CUSTOM_VNIC_TYPE_DIRECT CUSTOM_VNIC_TYPE_MACVTAP CUSTOM_VNIC_TYPE_DIRECT_PHYSICAL		
uuid	91d8208f-0c3a-59f5-97c6-3c79c15d295d		

inventories	{}
name	demo0:NIC Switch agent
traits	[]
uuid	8ae5944a-01fa-53e0-98dc-067a33192302

parent

parent

name	demo0
uuid	c29071e1-768b-429d-ad23-790e624bacf2

parent

inventories	{}
name	demo0:Open vSwitch agent
traits	[]
uuid	4495c088-8ca7-567d-b930-19e16c6ac4d7

parent

inventories	NET_BW_EGRESS_KBPS	total	10000000
	NET_BW_INGRESS_KBPS	total	10000000
name	demo0:Open vSwitch agent:br-physnet0		
traits	CUSTOM_PHYSNET_PHYSNET0 CUSTOM_VNIC_TYPE_NORMAL		
uuid	4170cc1f-f9ed-576c-874d-0da40ef01576		

parent

inventories	NET_BW_EGRESS_KBPS	total	1000000
name	demo0:Open vSwitch agent:br-ex		
traits	CUSTOM_PHYSNET_PUBLIC CUSTOM_VNIC_TYPE_NORMAL		
uuid	71d14a10-3772-5ddd-8a80-12acef8c0d76		

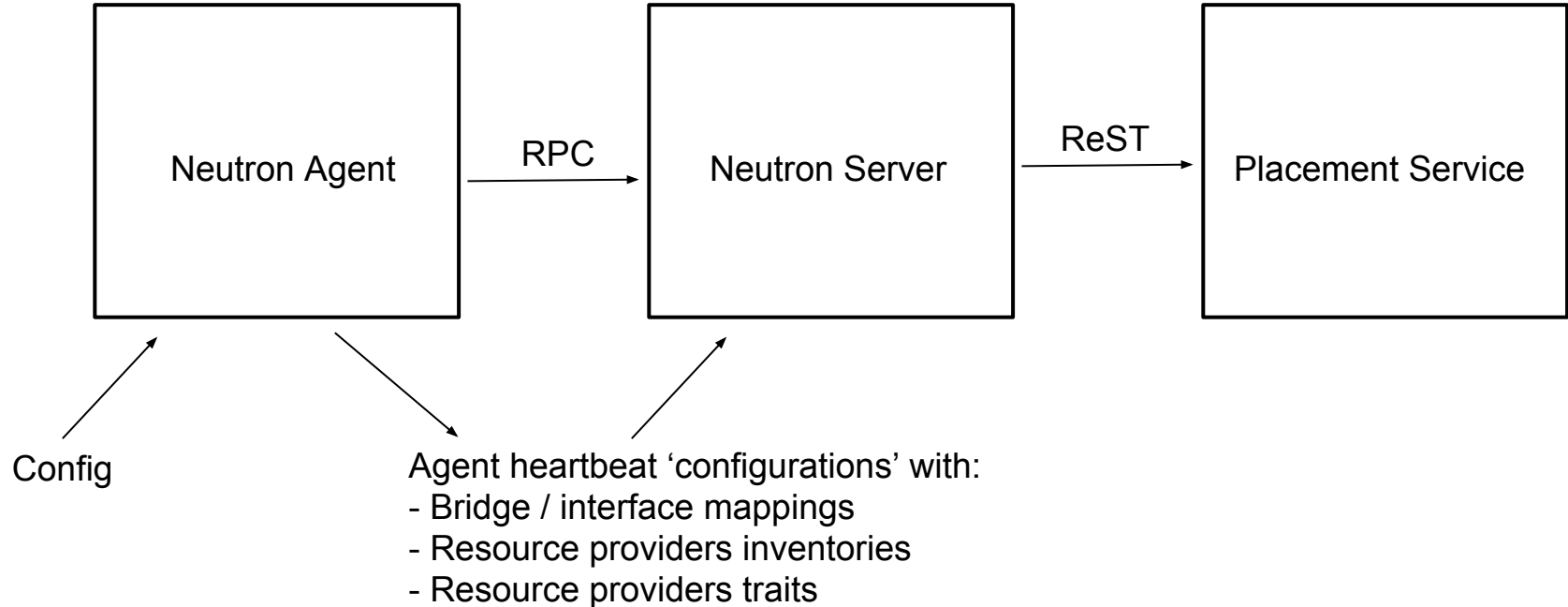
Physical device resource provider

name	demo0:NIC Switch agent
------	------------------------

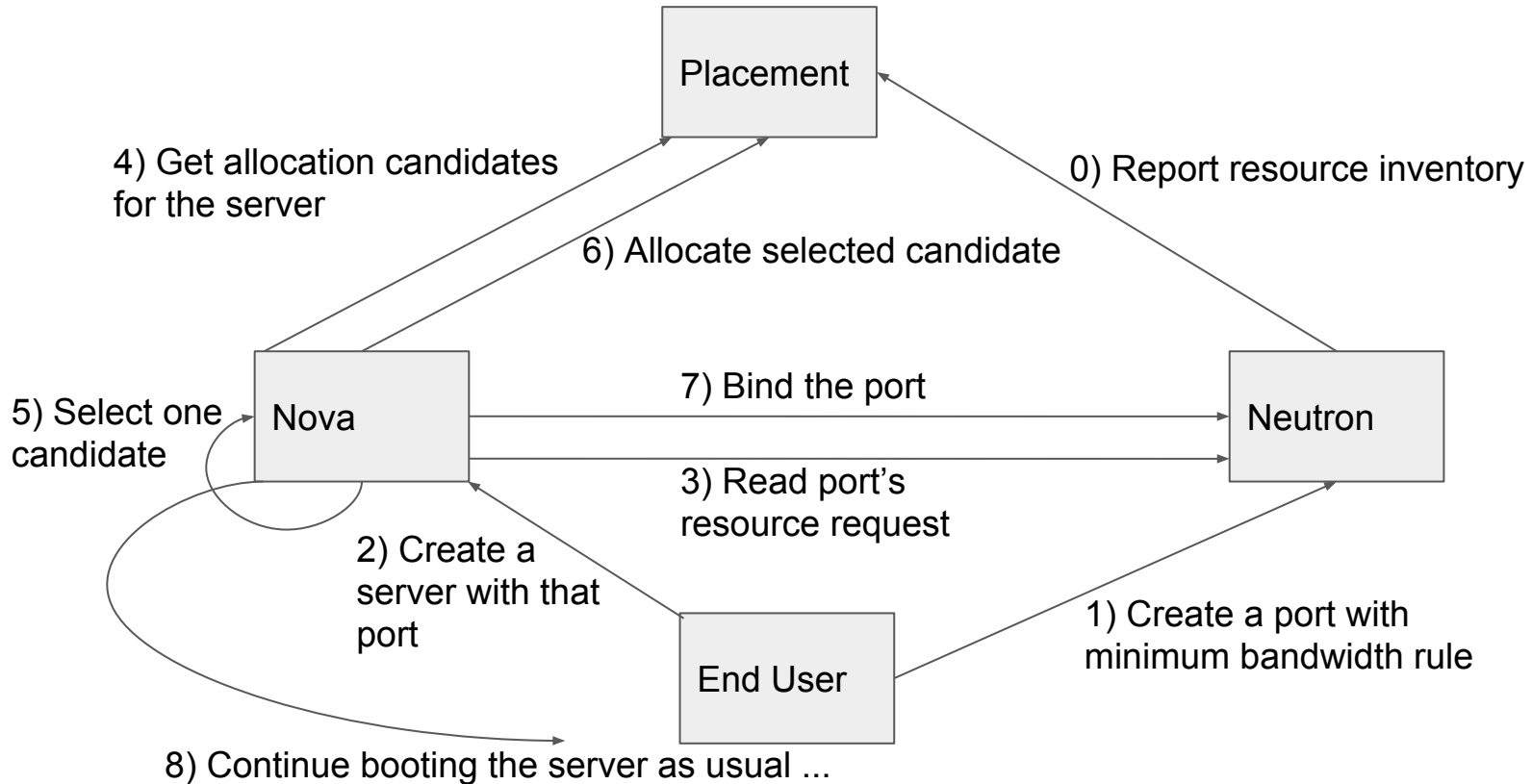
← parent

inventories	NET_BW_EGRESS_KBPS	total	40000000
	NET_BW_INGRESS_KBPS	total	40000000
name	demo0:NIC Switch agent:ens5		
traits	CUSTOM_PHYSNET_PHYSNET0 CUSTOM_VNIC_TYPE_DIRECT CUSTOM_VNIC_TYPE_MACVTAP CUSTOM_VNIC_TYPE_DIRECT_PHYSICAL		
uuid	91d8208f-0c3a-59f5-97c6-3c79c15d295d		

Reporting bandwidth inventories to Placement



Creating server with guaranteed bandwidth



Neutron communicates to Nova port's minimum bandwidth requirements at #3)

```
{
  "port": {
    "status": "ACTIVE",
    "name": "port0-min-6gbps",
    ...
    "device_id": "a62d9937-b9a2-41d0-862d-c799bac80cd6",
    "resource_request": {
      "resources": {
        "NET_BANDWIDTH_INGRESS_KILOBITS_PER_SECOND": 6000,
        "NET_BANDWIDTH_EGRESS_KILOBITS_PER_SECOND": 6000,
      },
      "required": [
        "CUSTOM_PHYSNET_PHYSNET0",
        "CUSTOM_VNIC_TYPE_NORMAL"
      ]
    }
  }
}
```

Nova communicates to Neutron the selected host and backend

```
{
  "port": {
    "binding:host_id": "allinone",
    "binding:profile": {
      "allocation": "<rp uuid>"
    }
  }
}
```

Demo

Demo overview

- **API extensions and microversions**
- **Configuration**
- 3 Scenarios
 - **Boot** with minimum bandwidth
 - **Keep under** and ...
 - ... **go over** total available

Neutron API

```
$ openstack extension show qos-bw-minimum-ingress
```

Field	Value
alias	qos-bw-minimum-ingress
description	Allow to configure QoS minimum bandwidth rule with ingress direction.
id	qos-bw-minimum-ingress
links	[]
location	None
name	Ingress direction for QoS minimum bandwidth rule
updated_at	2018-07-09T10:00:00-00:00

```
$ openstack extension show port-resource-request
```

Field	Value
alias	port-resource-request
description	Expose resource request to Port
id	port-resource-request
links	[]
location	None
name	Port Resource Request
updated_at	2018-05-08T10:00:00-00:00

Placement API

```
# minimum Placement microversion: 1.29
```

```
$ export TOKEN="$( openstack token issue -f value -c id )"
$ curl \
  --silent \
  --header "Accept: application/json" \
  --header "Content-Type: application/json" \
  --header "X-Auth-Token: $TOKEN" \
  --header "OpenStack-API-Version: placement latest" \
  'http://127.0.0.1/placement/' \
  | jq .versions[0].max_version
"1.30"
```

Neutron config

```
/etc/neutron/neutron.conf:
```

```
[placement]
```

```
project_domain_name = Default
```

```
project_name = service
```

```
user_domain_name = Default
```

```
password = devstack
```

```
username = nova
```

```
auth_url = http://127.0.0.1/identity
```

```
auth_type = password
```

```
/etc/neutron/plugins/ml2/ml2_conf.ini:
```

```
[ovs]
```

```
bridge_mappings = public:br-ex,physnet0:br-physnet0
```

```
resource_provider_bandwidths = br-physnet0:10000000:10000000,br-ex:1000000:
```

```
[ovs_driver]
```

```
vnic_type_blacklist = direct
```

```
/etc/neutron/plugins/ml2/sriov_agent.ini:
```

```
[sriov_nic]
```

```
physical_device_mappings = physnet0:ens5
```

```
exclude_devices =
```

```
resource_provider_bandwidths = ens5:40000000:40000000
```

Resource provider and inventory

```
$ openstack --os-placement-api-version 1.17 resource provider list
```

uuid	name	generation	root_provider_uuid	parent_provider_uuid
c29071e1-768b-429d-ad23-790e624bacf2	demo0	2	c29071e1-768b-429d-ad23-790e624bacf2	None
8ae5944a-01fa-53e0-98dc-067a33192302	demo0:NIC Switch agent	0	c29071e1-768b-429d-ad23-790e624bacf2	c29071e1-768b-429d-ad23-790e624bacf2
91d8208f-0c3a-59f5-97c6-3c79c15d295d	demo0:NIC Switch agent:ens5	2	c29071e1-768b-429d-ad23-790e624bacf2	8ae5944a-01fa-53e0-98dc-067a33192302
4495c088-8ca7-567d-b930-19e16c6ac4d7	demo0:Open vSwitch agent	0	c29071e1-768b-429d-ad23-790e624bacf2	c29071e1-768b-429d-ad23-790e624bacf2
4170cc1f-f9ed-576c-874d-0da40ef01576	demo0:Open vSwitch agent:br-physnet0	2	c29071e1-768b-429d-ad23-790e624bacf2	4495c088-8ca7-567d-b930-19e16c6ac4d7
71d14a10-3772-5ddd-8a80-12acef8c0d76	demo0:Open vSwitch agent:br-ex	2	c29071e1-768b-429d-ad23-790e624bacf2	4495c088-8ca7-567d-b930-19e16c6ac4d7

```
$ openstack --os-placement-api-version 1.17 resource provider list | awk '/demo0:Open vSwitch agent:br-physnet0/ { print $2 }' | xargs -r -n1 openstack resource provider inventory list
```

resource_class	allocation_ratio	max_unit	reserved	step_size	min_unit	total
NET_BANDWIDTH_EGRESS_KILOBITS_PER_SECOND	1.0	2147483647	0	1	1	10000000
NET_BANDWIDTH_INGRESS_KILOBITS_PER_SECOND	1.0	2147483647	0	1	1	10000000

Network, QoS policies and rules, ports

```
$ openstack network show net0
```

Field	Value
id	4f5041fe-942d-4e79-aed9-8022f20fa5fc
name	net0
provider:network_type	vlan
provider:physical_network	physnet0
provider:segmentation_id	100
qos_policy_id	None
subnets	609c84a1-1f37-48a5-b079-61a599b74f79
...	

```
$ openstack network qos rule list qos-policy-min-6gbps
```

ID	QoS Policy ID	Type	Max Kbps	Max Burst Kbits	Min Kbps	DSCP mark	Direction
4e442232-3639-443b-9cc0-19f3cabe6b1c	260d240a-5a3c-41dc-9cd0-8a3e5670368b	minimum_bandwidth			6000000		egress
84c55491-ebcb-49eb-a3f7-5cfe552e04fb	260d240a-5a3c-41dc-9cd0-8a3e5670368b	minimum_bandwidth			6000000		ingress

```
$ openstack network qos rule list qos-policy-min-1gbps
```

ID	QoS Policy ID	Type	Max Kbps	Max Burst Kbits	Min Kbps	DSCP mark	Direction
a9b9baac-1303-425c-a24d-c1a4cfcf396f	5ecd692a-01ea-4a7f-ae07-e619e9217890	minimum_bandwidth			1000000		egress
655206f3-b6c1-4ff1-994f-b229eb5b7f04	5ecd692a-01ea-4a7f-ae07-e619e9217890	minimum_bandwidth			1000000		ingress

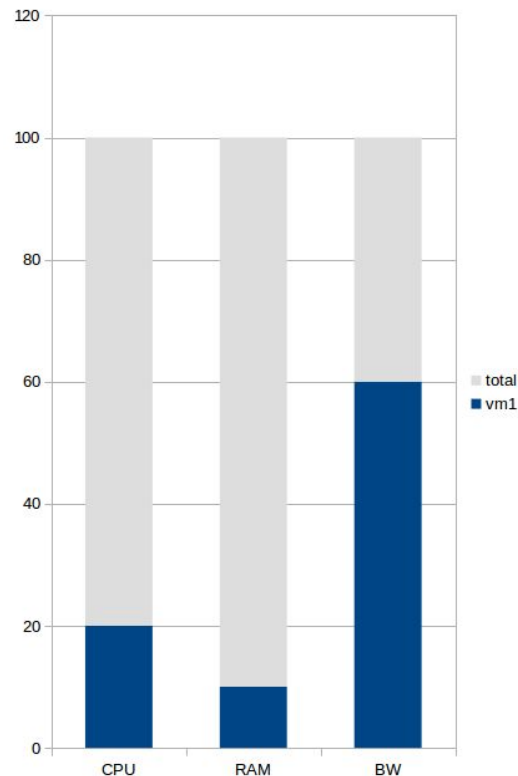
```
$ openstack port list
```

ID	Name	MAC Address	Fixed IP Addresses	Status
5ee51ab6-ea8e-47c2-96a5-da6c1d73fb54	port0-min-6gbps	fa:16:3e:14:6e:27	ip_address='10.0.4.19', subnet_id='609c84a1-1f37-48a5-b079-61a599b74f79'	DOWN
98f23745-fb52-417e-afe8-386b8e5ae526	port1-min-1gbps	fa:16:3e:64:fb:fb	ip_address='10.0.4.7', subnet_id='609c84a1-1f37-48a5-b079-61a599b74f79'	DOWN
acc12e37-3460-4aee-995d-925c47ed2984	port2-min-6gbps	fa:16:3e:c1:11:d6	ip_address='10.0.4.5', subnet_id='609c84a1-1f37-48a5-b079-61a599b74f79'	DOWN
...				

Boot vm1

```
$ openstack server create --flavor cirros256 --image cirros-0.3.5-x86_64-disk --nic port-id=port0-min-6gbps --wait vm1
```

Field	Value
OS-DCF:diskConfig	MANUAL
OS-EXT-AZ:availability_zone	nova
OS-EXT-SRV-ATTR:host	demo0
OS-EXT-SRV-ATTR:hypervisor_hostname	demo0
OS-EXT-SRV-ATTR:instance_name	instance-00000003
OS-EXT-STS:power_state	Running
OS-EXT-STS:task_state	None
OS-EXT-STS:vm_state	active
OS-SRV-USG:launched_at	2018-10-31T12:09:57.000000
OS-SRV-USG:terminated_at	None
accessIPv4	
accessIPv6	
addresses	net0=10.0.4.19
adminPass	3ge4LCeqsQwv
config_drive	
created	2018-10-31T12:09:50Z
flavor	cirros256 (c1)
hostId	eef38067a2f3d47ed5a6a81c1794fee501346033949692e9d4d291b1
id	a62d9937-b9a2-41d0-862d-c799bac80cd6
image	cirros-0.3.5-x86_64-disk (8a353fce-d9a7-4805-96d5-b915c59bff33)
key_name	None
name	vm1
progress	0
project_id	8f36ef0a54ae4cd1a4fd64a7dc6deb13
properties	
security_groups	name='default'
status	ACTIVE
updated	2018-10-31T12:09:58Z
user_id	8a243f7135ef4d06b01ee870a8f83d5a
volumes_attached	



Resource allocation in Placement

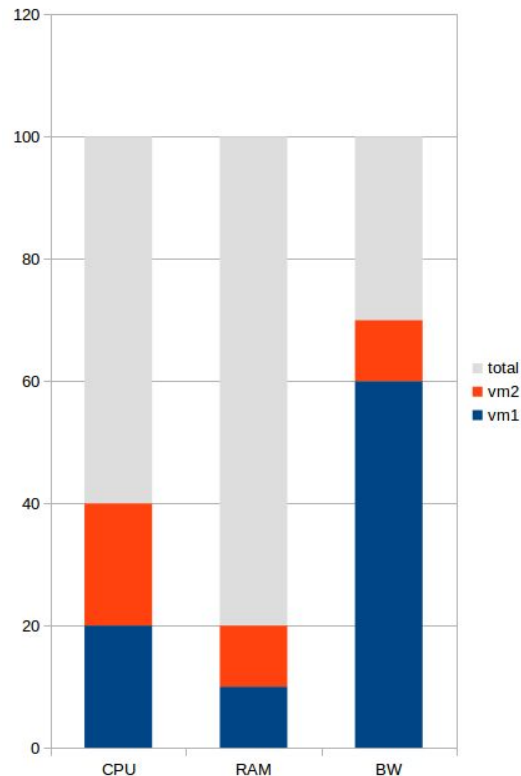
```
$ openstack --os-placement-api-version 1.17 resource provider list | awk '/demo0:Open vSwitch agent:br-physnet0/ { print $2 }' | xargs -r -n1 openstack resource provider show --allocations
```

Field	Value
uuid	4170cc1f-f9ed-576c-874d-0da40ef01576
name	demo0:Open vSwitch agent:br-physnet0
generation	5
allocations	{u'a62d9937-b9a2-41d0-862d-c799bac80cd6': {u'resources': {u'NET_BANDWIDTH_EGRESS_KILOBITS_PER_SECOND': 6000000, u'NET_BANDWIDTH_INGRESS_KILOBITS_PER_SECOND': 6000000}}}

Boot vm1+vm2

```
$ openstack server create --flavor cirros256 --image cirros-0.3.5-x86_64-disk --nic port-id=port1-min-1gbps --wait vm2
```

Field	Value
OS-DCF:diskConfig	MANUAL
OS-EXT-AZ:availability_zone	nova
OS-EXT-SRV-ATTR:host	demo0
OS-EXT-SRV-ATTR:hypervisor_hostname	demo0
OS-EXT-SRV-ATTR:instance_name	instance-00000004
OS-EXT-STS:power_state	Running
OS-EXT-STS:task_state	None
OS-EXT-STS:vm_state	active
OS-SRV-USG:launched_at	2018-10-31T12:13:44.000000
OS-SRV-USG:terminated_at	None
accessIPv4	
accessIPv6	
addresses	net0=10.0.4.7
adminPass	fumwv2C9cU3T
config_drive	
created	2018-10-31T12:13:36Z
flavor	cirros256 (c1)
hostId	eef38067a2f3d47ed5a6a81c1794fee501346033949692e9d4d291b1
id	8ab682b5-f627-483a-ad02-6878995d9a69
image	cirros-0.3.5-x86_64-disk (8a353fce-d9a7-4805-96d5-b915c59bfff33)
key_name	None
name	vm2
progress	0
project_id	8f36ef0a54ae4cd1a4fd64a7dc6deb13
properties	
security_groups	name='default'
status	ACTIVE
updated	2018-10-31T12:13:44Z
user_id	8a243f7135ef4d06b01ee870a8f83d5a
volumes_attached	



Resource allocation in Placement

```
$ openstack --os-placement-api-version 1.17 resource provider list \  
  | awk '/demo0:Open vSwitch agent:br-physnet0/ { print $2 }' \  
  | xargs -r -n1 openstack resource provider usage show
```

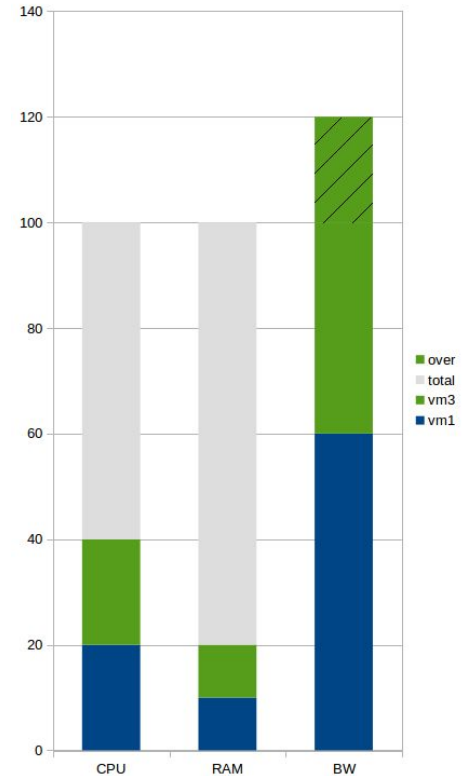
```
+-----+-----+  
| resource_class                | usage |  
+-----+-----+  
| NET_BANDWIDTH_EGRESS_KILOBITS_PER_SECOND | 7000000 |  
| NET_BANDWIDTH_INGRESS_KILOBITS_PER_SECOND | 7000000 |  
+-----+-----+
```

Boot vm1+vm3

```
$ openstack server delete vm2
```

```
$ openstack server create --flavor cirros256 --image cirros-0.3.5-x86_64-disk --nic port-id=port2-min-6gbps --wait vm3  
Error creating server: vm3  
Error creating server
```

```
# No valid host was found.
```



Where we are with the implementation

- Both Neutron and Nova specifications are approved
- Implementation is ongoing both in Neutron and Nova
 - <https://review.openstack.org/#/q/status:open+project:openstack/nova+topic:bp/bandwidth-resource-provider>
 - <https://review.openstack.org/#/q/status:open+project:openstack/neutron+topic:minimum-bandwidth-allocation-placement-api>
- The plan is to have this feature usable at Stein and continue developing it further

References

- This presentation: http://bit.ly/min_bw
- Neutron specification:
<https://specs.openstack.org/openstack/neutron-specs/specs/rocky/minimum-bandwidth-allocation-placement-api.html>
- Nova specification:
<https://specs.openstack.org/openstack/nova-specs/specs/stein/approved/bandwidth-resource-provider.html>
- Blog post:
<https://rubasov.github.io/2018/09/21/openstack-qos-min-bw-demo.html>
- Placement diagrams in this presentation are generated with:
<https://github.com/gibizer/osc-placement-tree>