목차

Ironic 베어메탈 관리 서비스의 네트워크 모델 3
***Neutron*에서의 베어메탈 네트워킹** 3
핵심 네트워크 유형 3
Cleaning Network 3
Provisioning Network 3
Tenant Network 3
네트워크 토폴로지 3
추상 네트워크 토폴로지 다이어그램 5
Neutron 구현 6
***Neutron*이 베어메탈 노드 네트워크 토폴로지를 이해하는 방식** 6
물리 스위치 제어 7
베어메탈 노드의 네트워크 수명 주기 7
배포 네트워크와 테넌트 네트워크 간 전환 과정 9
SDN 기반 네트워크 전환 흐름 11
참조링크 13

Ironic 베어메탈 관리 서비스의 네트워크 모델

Neutron에서의 베어메탈 네트워킹

Ironic은 네트워크 구성을 위해 OpenStack의 네트워킹 API인 Neutron을 사용합니다. “ 베어메탈 ” 배포는 가상 머신(VM)과는 약간 다르며, Ironic은 Neutron ml2 구현으로부터 몇 가지 추가 요구 사항이 있습니다.

핵심 네트워크 유형

Cleaning Network

Cleaning Network: 이 네트워크는 베어메탈 서버를 정리(clean)하고 “ 베어메탈 ” 노드가 새로운 작업 부하를 처리할 준비가 되었는지 확인하는 데 사용됩니다. 이 네트워크는 테넌트 VLAN 범위와 분리하기 위해 provider-VLAN 네트워크로 생성하는 것이 권장됩니다.

Provisioning Network

Provisioning Network: 노드의 정기적인 관리(해체, 재부팅, PXE 부팅 등)를 위해 사용되는 네트워크입니다. 이 네트워크 역시 클리닝 네트워크와 같은 이유로 provider-VLAN 네트워크로 생성하는 것이 권장됩니다. (운영자는 Provisioning과 Cleaning을 같은 네트워크에서 사용할 수 있지만, Ironic은 클린 노드와 정리가 필요한 노드를 분리하기 위해 두 네트워크 유형을 따로 정의할 수 있도록 합니다.)

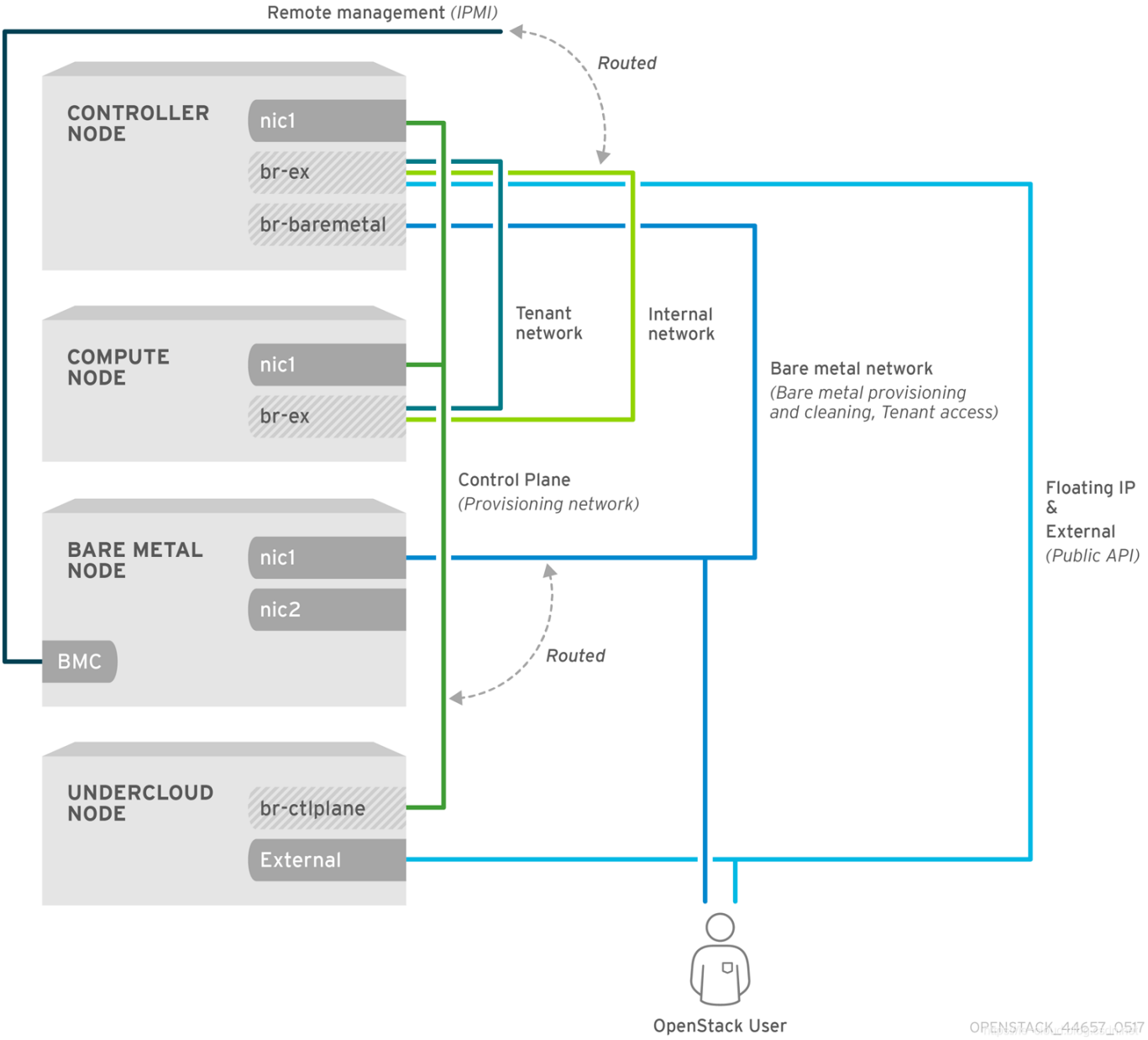
- 관리자가 Neutron 네트워크로 생성
- Ironic 컴퓨트 노드는 이 VLAN에 연결됨
- 각 베어메탈 노드는 배포 시에만 이 VLAN에 연결됨

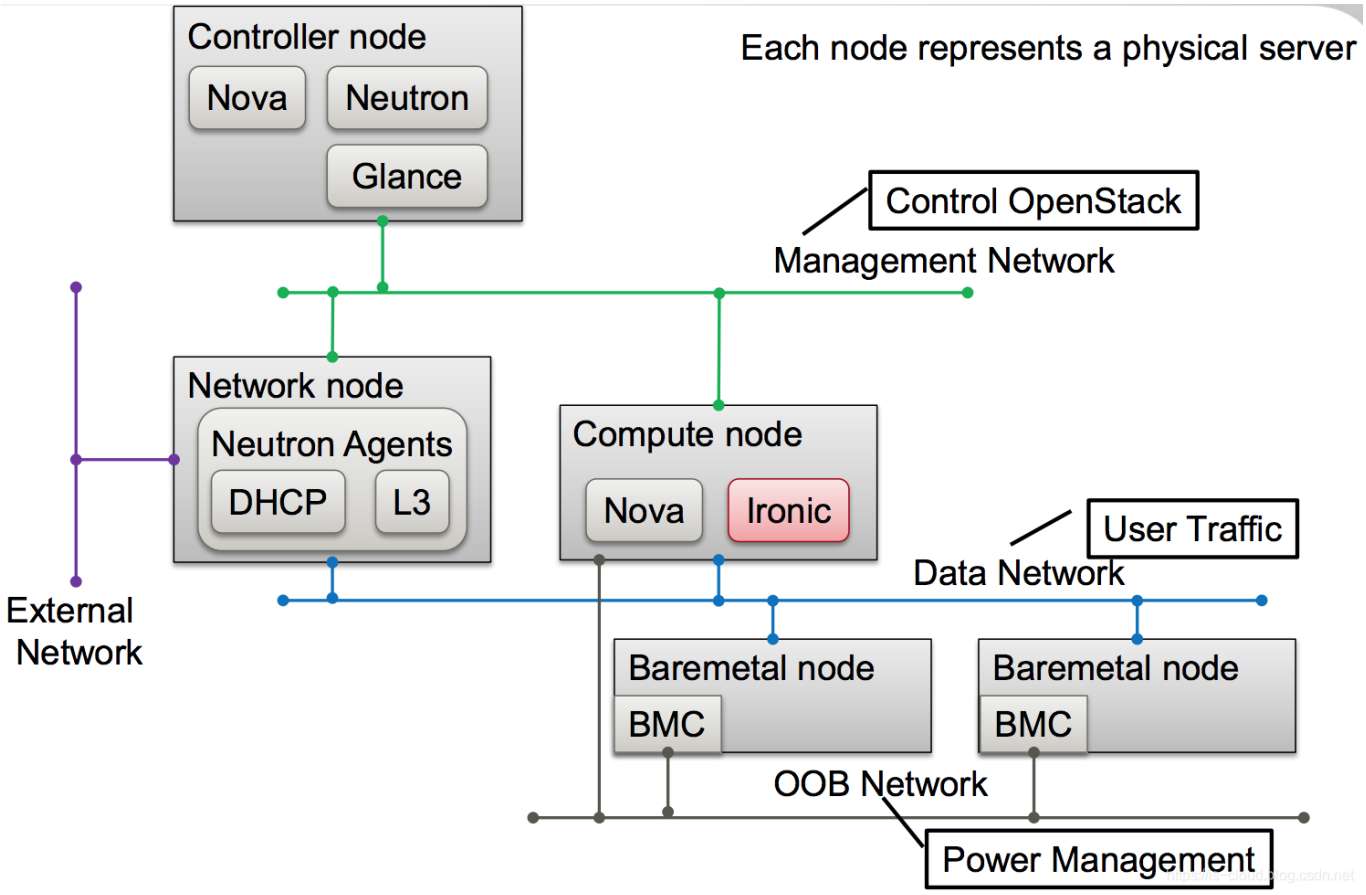
Tenant Network

Tenant Network: “ 베어메탈 ” 에 접근하거나 다른 목적으로 사용되는 네트워크로, 클라우드의 다른 네트워크처럼 관리되어야 합니다. 베어메탈 노드가 테넌트 네트워크에 연결될 때는 보안상의 이유로 프로비저닝 네트워크에 연결되지 않아야 합니다. (모든 베어메탈 서버에서 동일한 프로비저닝 네트워크를 사용하므로, 이는 격리 요구 사항을 위반할 수 있습니다.)

- 테넌트 사용자가 Neutron 네트워크로 생성
- 테넌트 내의 베어메탈 노드는 배포 후 이 유형의 VLAN에 연결됨

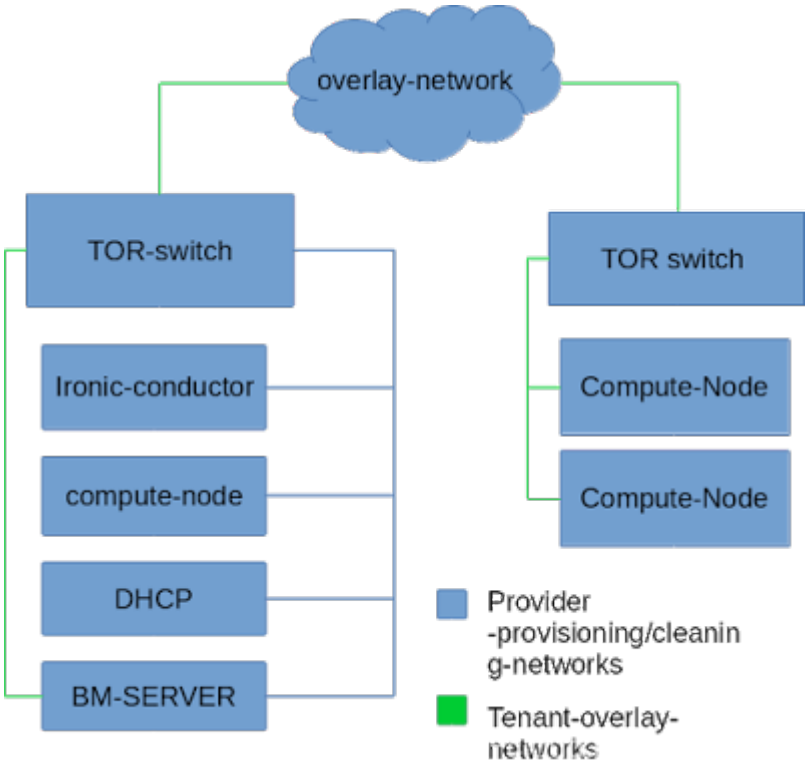
네트워크 토폴로지


Remote management (IPMI)
Routed
CONTROLLER NODE
nic1
br-ex
br-baremetal
COMPUTE NODE
nic1
br-ex
Tenant network
Internal network
Bare metal network
(Bare metal provisioning and cleaning, Tenant access)
Control Plane
(Provisioning network)
Floating IP & External
(Public API)
BARE METAL NODE
nic1
nic2
BMC
Routed
UNDERCLOUD NODE
br-ctlplane
External
OpenStack User
OPENSTACK_44657_0517



- **Management Network**: OpenStack 관리 네트워크.
- **External Network**: 외부 네트워크.
- **Data Network**: 비즈니스 네트워크로, 비즈니스 트래픽을 처리하며 OpenStack의 기반 네트워크로 사용됩니다.
- **OOB (Out-of-Band) Network**: 대역 외 네트워크, 즉 IPMI 네트워크.

추상 네트워크 토폴로지 다이어그램



- **Tenant Network**: “베어메탈” 노드에 동적으로 연결 및 분리될 수 있는 네트워크입니다.
- **Provider Networks**: 클리닝 및 프로비저닝을 위한 네트워크, 그리고 기타 필요에 따라 사용되는 네트워크입니다.
- **Ironic Conductor**: 실제로 “베어메탈” 서버를 제어하는 Ironic의 소프트웨어 구성 요소로, PXE 부팅을 위한 TFTP 서버도 포함합니다.
- **DHCP Server**: “베어메탈” 서버에 IP 주소를 할당하고, PXE 부팅 파라미터를 지원하는 DHCP 서버입니다.
- **Top of Rack Switch**: “베어메탈” 서버가 물리적으로 연결된 스위치로, 다른 모든 구성 요소(컴퓨트 노드, Ironic Conductor 노드 등)와 함께 연결됩니다.
- **Bare-metal Server**: “베어메탈” 서버 자체입니다.

Neutron 구현

- **Supporting Port-Groups**: 베어메탈은 종종 여러 물리적 포트를 하나의 논리적 포트로 처리해야 할 때가 있습니다(예: BOND/LAG). 이러한 포트 그룹은 Neutron에 의해 관리되어야 합니다.
- **PXE 부팅 지원(DHCP 사용)**: 베어메탈 서버를 부팅하는 가장 일반적인 방법은 PXE 부팅입니다. PXE 부팅 절차는 DHCP를 사용하여 부팅 파일 이름과 TFTP 서버 주소를 가져옵니다. Ironic은 이 파라미터 값을 Neutron에 전달합니다(Neutron의 `extra_dhcp_opt`를 사용하여). Neutron의 DHCP 서버 구현은 이 파라미터를 사용하여 PXE DHCP 요청에 응답해야 합니다.

Neutron이 베어메탈 노드 네트워크 토폴로지를 이해하는 방식

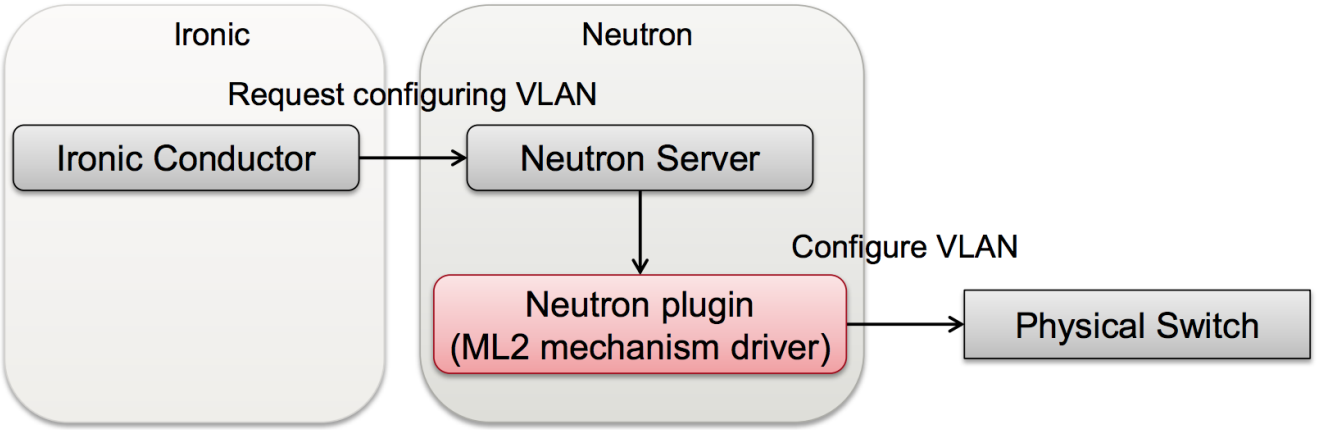
Neutron 포트 구성: Ironic은 Neutron에 “베어메탈” 포트를 알리기 위해 자체 메커니즘을 사용하여 하드웨어를 검사하고, 해당 정보를 Neutron 포트 구성의 일부로 전달합니다. 이를 위해 Neutron 로컬 포트(lport) 사양에 두 가지 새로운 필드가 도입되었습니다:

- **local_link_information**: 이 필드는 lport의 binding-profile에 위치하며, Neutron에 포트가 TOR 스위치에 어떻게 연결되어 있는지를 알리는 데 사용됩니다. 이 필드는 세 가지 매개변수를 포함합니다:
  - **switch_id**: 포트가 연결된 스위치의 식별자입니다. 스위치의 MAC 주소 또는 OpenFlow 기반의 datapath_id일 수 있습니다.
  - **port_id**: 스위치의 물리적 포트 식별자입니다.
  - **switch_info**: 스위치에 대한 기타 정보(선택적 매개변수)입니다.
- **port-groups**: TOR에서 LAG/BOND를 구성하기 위한 매개변수 목록입니다.

Neutron 메커니즘 드라이버는 이 정보를 사용하여 lport를 바인딩할 때 적절히 처리해야 합니다.

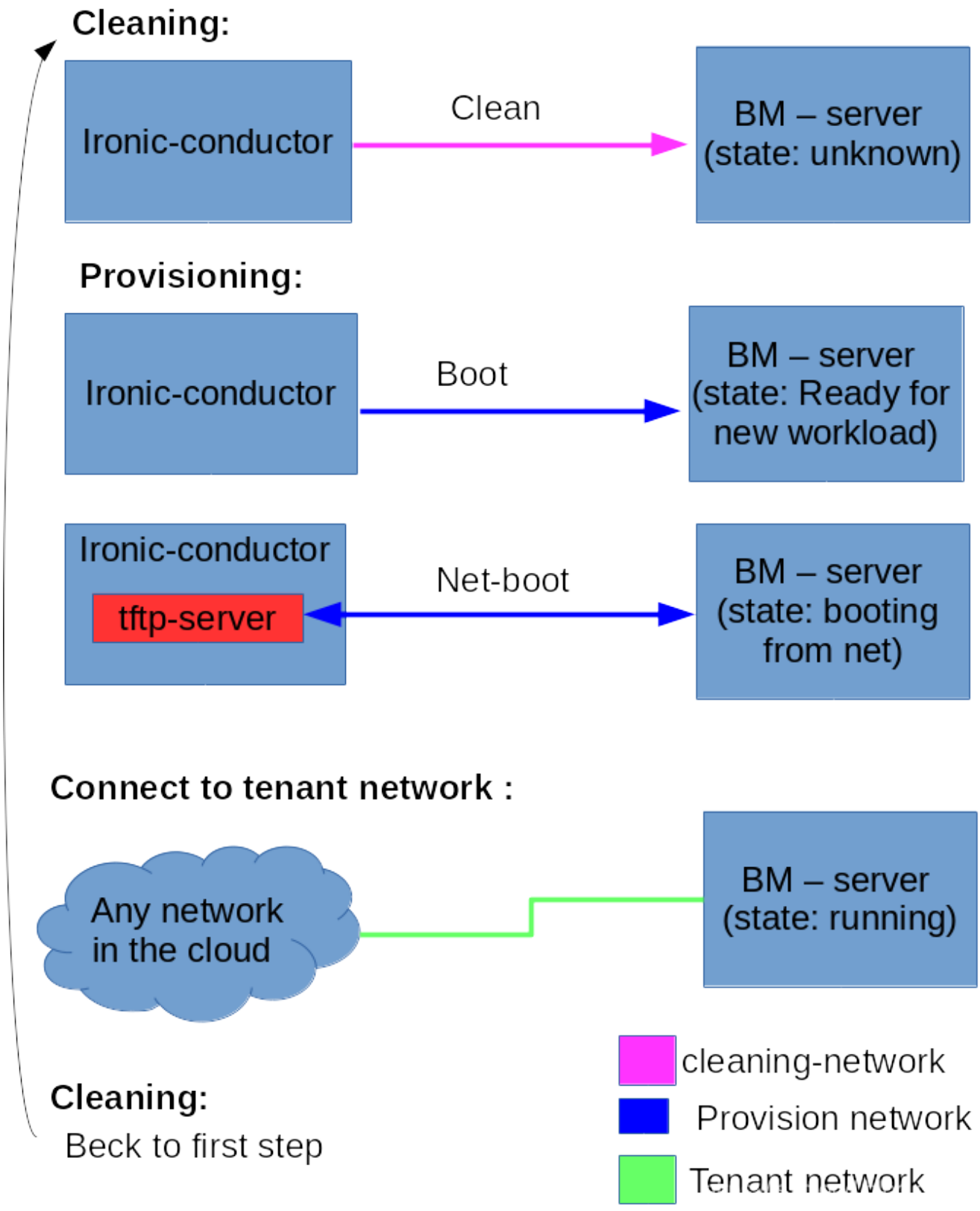
DHCP 구성: Ironic은 Neutron 포트에서 `extra_dhcp_option` 속성을 사용하여 DHCP가 PXE 부팅을 지원하도록 구성합니다(예: 부팅 파일 이름과 TFTP 서버 주소와 같은 DHCP 옵션). Neutron ML2 드라이버는 DHCP 서버가 요청 시 이 값을 응답하도록 구성해야 합니다.

물리 스위치 제어



- Neutron 플러그인을 통해 스위치를 제어하고 포트의 VLAN을 구성합니다.
- 이 플러그인은 ML2 메커니즘 드라이버로 구현됩니다.

베어메탈 노드의 네트워크 수명 주기

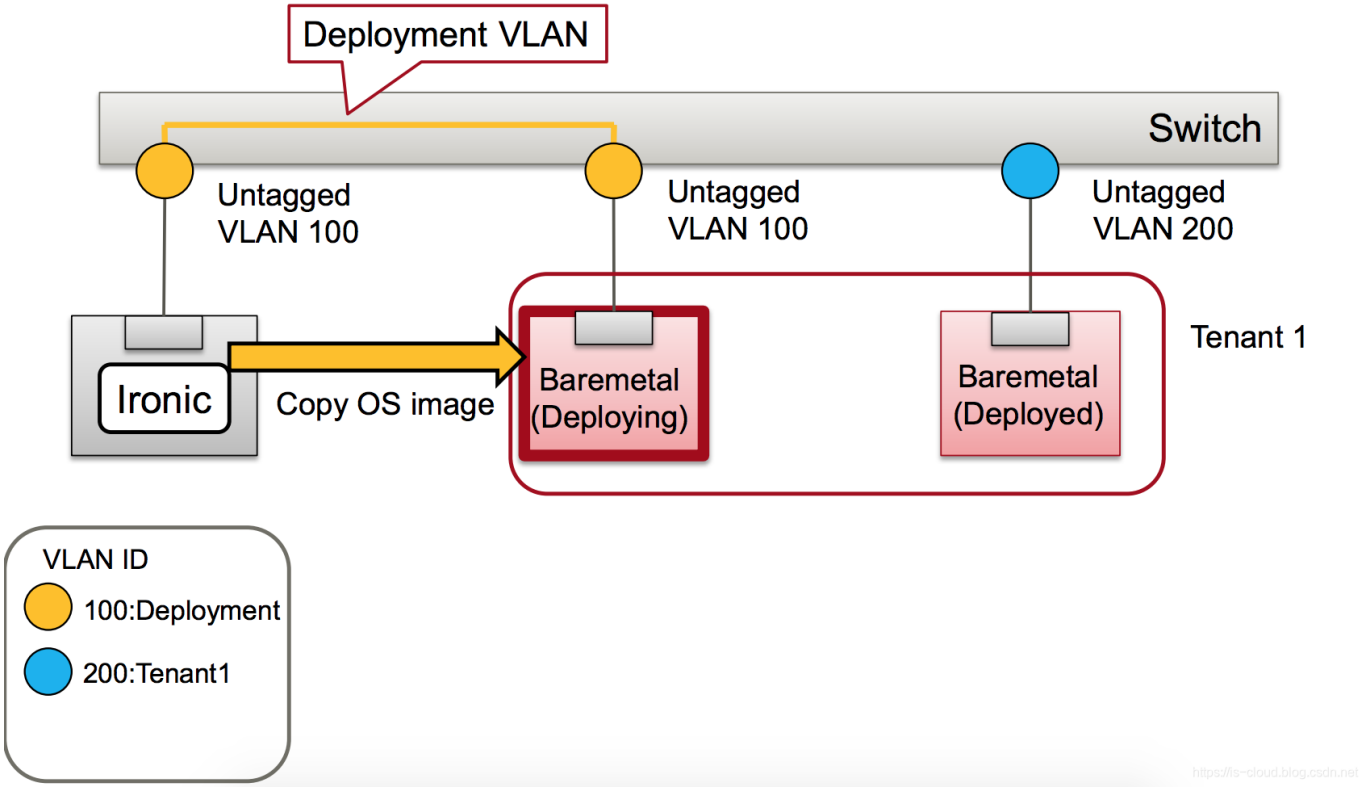


1. **Cleaning**: 노드를 새로운 작업에 준비시키는 과정(클리닝 네트워크 사용).
2. **Provisioning**: Ironic-conductor는 프로비저닝 네트워크에서 IPMI를 사용하여 머신을 시작하고, 원하는 이미지를 사용하여 PXE로 부팅합니다. PXE 부팅 프로세스는 다음 단계를 포함합니다(모든 단계는 프로비저닝 네트워크에서 수행됨):
    - DHCP를 사용해 TFTP 서버 주소를 얻음
    - TFTP 서버에서 부팅 파일 다운로드

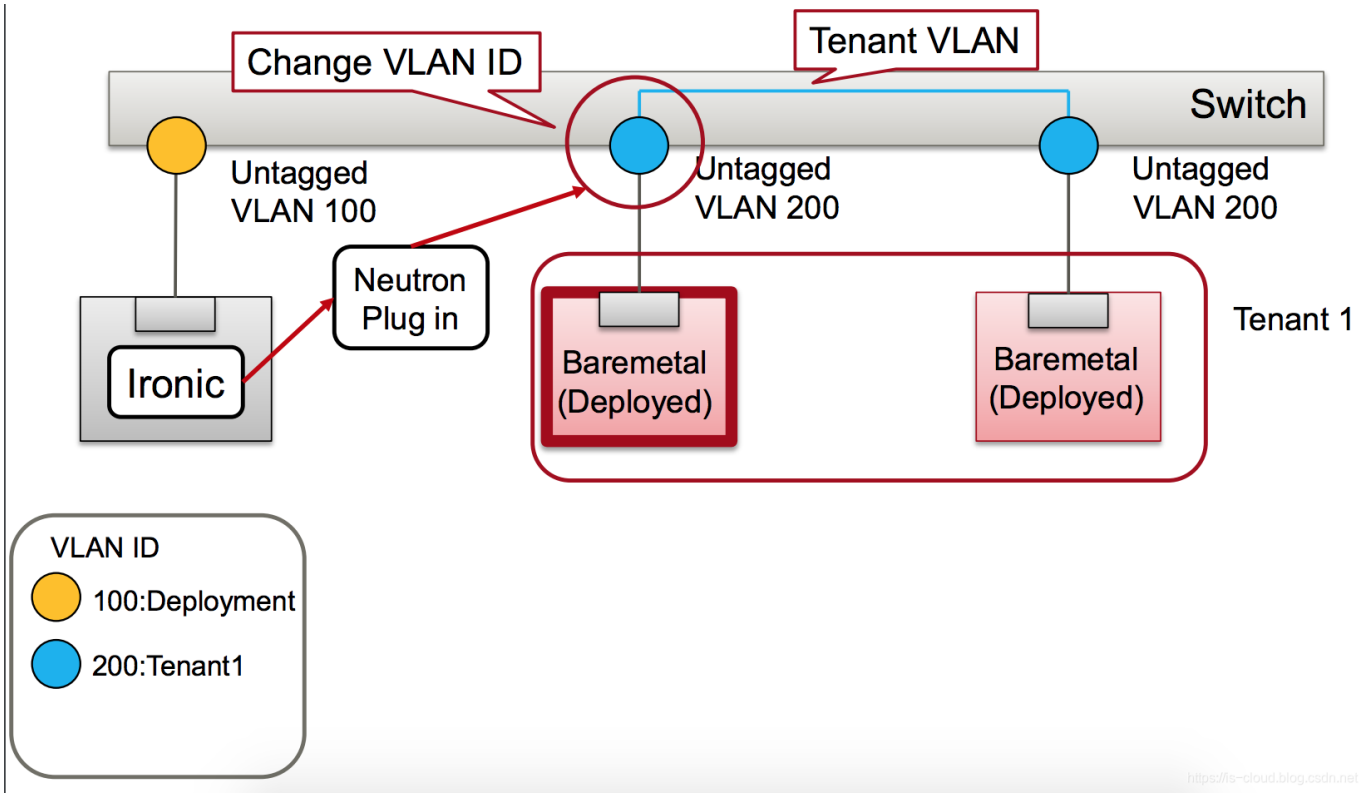
    - 다운로드한 파일로 부팅
3. **테넌트 네트워크에 연결**: 머신이 부팅되고 실행된 후에는 테넌트 네트워크에 연결되어 다른 가상 머신처럼 관리될 수 있습니다. 이 단계에서 “베어메탈” 서버의 트래픽은 배포의 다른 모든 구성 요소(e.g. VM, SNAT, DNAT 등)와 상호작용합니다.
    - Ironic은 프로비저닝 네트워크에서 사용된 물리 포트를 테넌트 네트워크로 바인딩할 수 있습니다. 이 경우 **베어메탈** 서버는 Ironic-conductor와의 연결 및 **베어메탈** 프로비저닝과의 연결을 상실하게 됩니다.
4. **Cleaning**: 1단계로 돌아갑니다…

배포 네트워크와 테넌트 네트워크 간 전환 과정

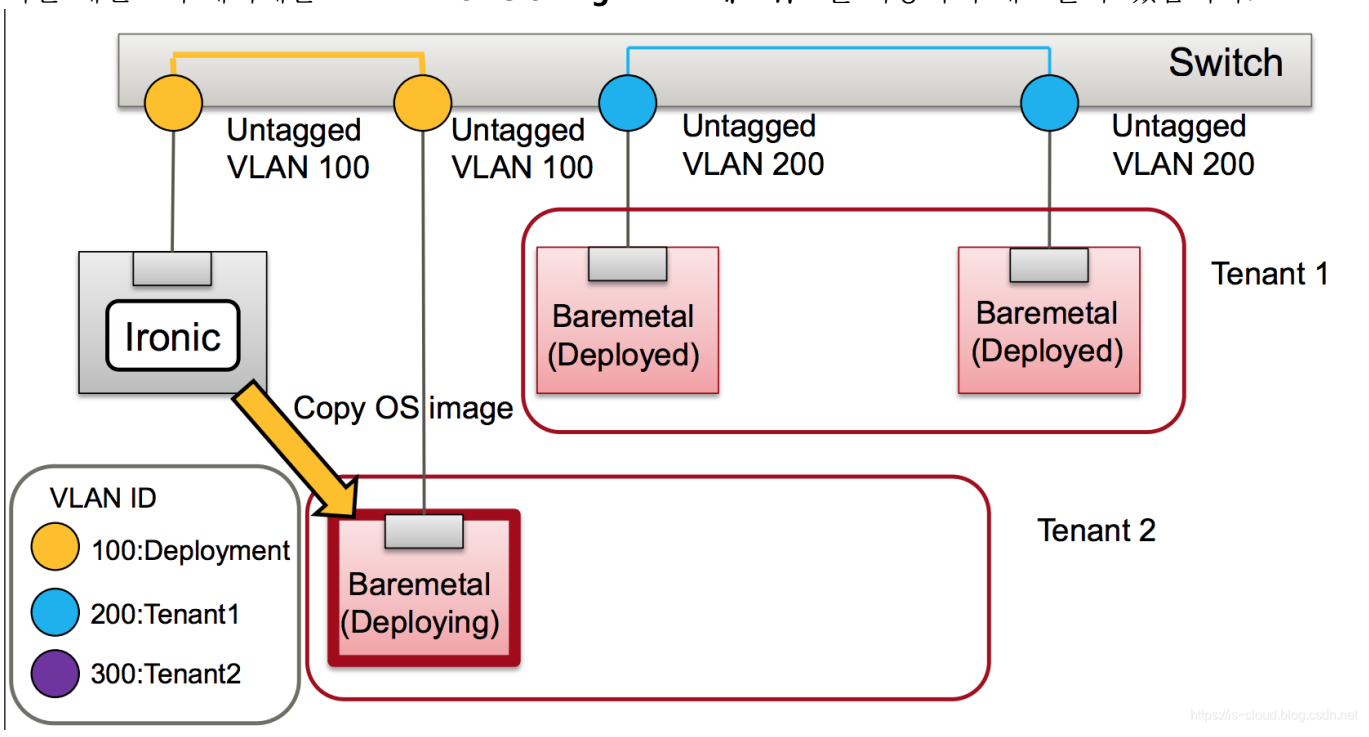
베어메탈 노드는 **Provisioning VLAN 네트워크**를 사용하여 배포됩니다.



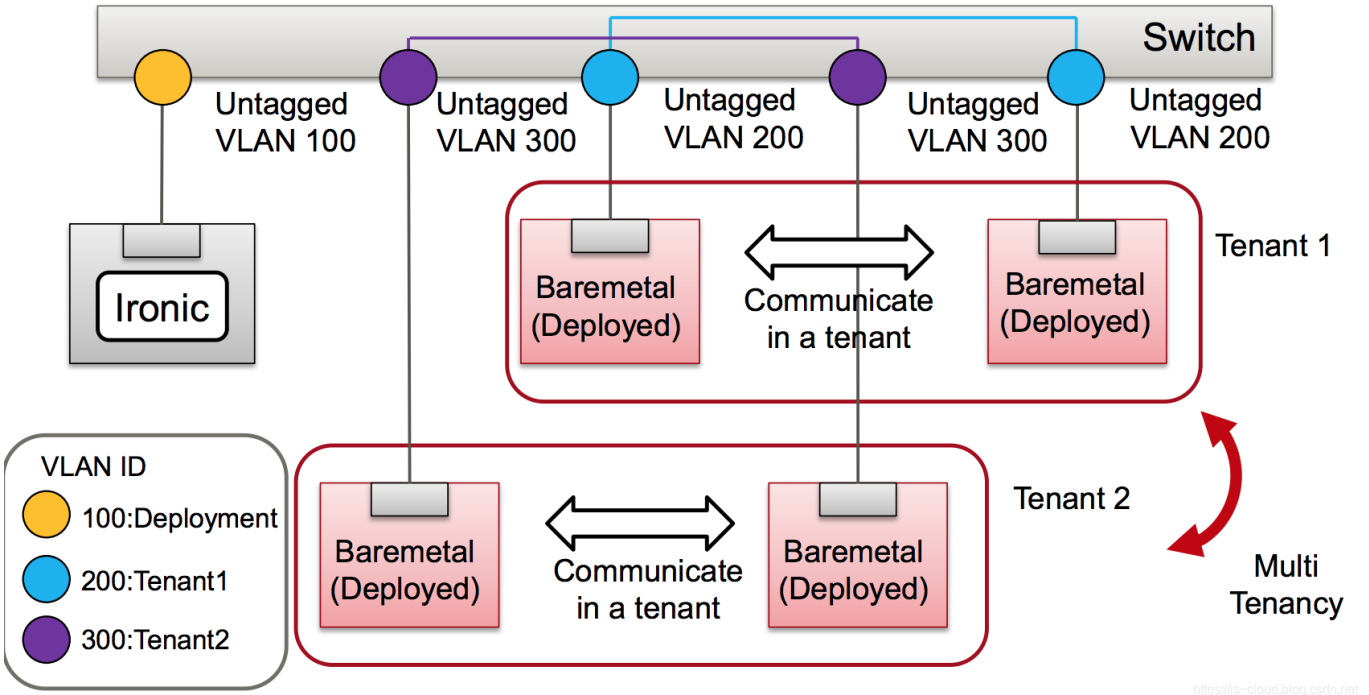
배포가 완료된 후, Ironic은 VLAN ID를 변경하여 베어메탈 노드가 테넌트 VLAN에 연결되도록 합니다.



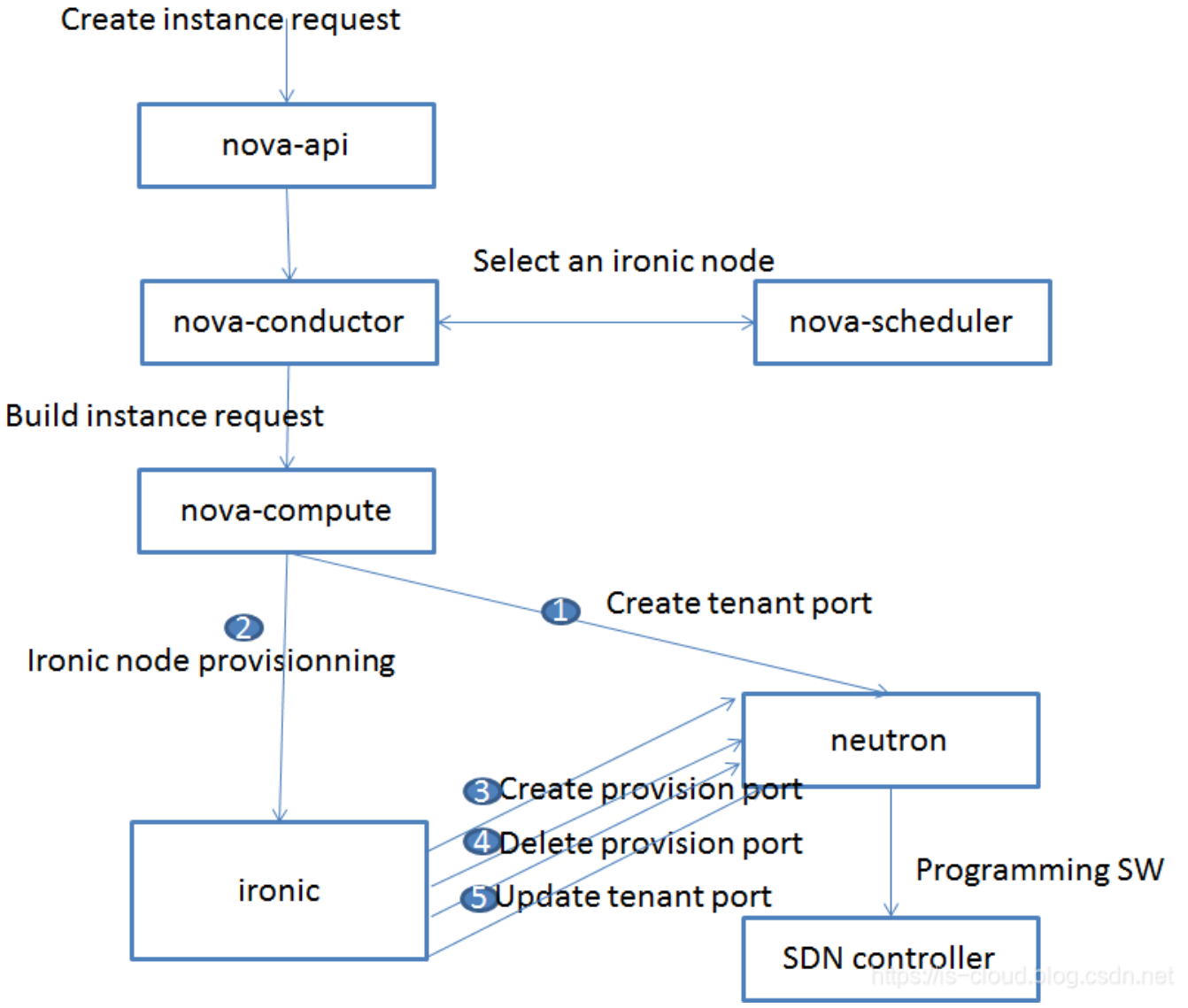
다른 테넌트의 베어메탈 노드도 **Provisioning VLAN 네트워크**를 사용하여 배포될 수 있습니다.



VLAN을 전환함으로써 Ironic은 모든 테넌트를 관리할 수 있습니다.



SDN 기반 네트워크 전환 흐름



참고: Tenant Port의 수는 사용자가 베어메탈 인스턴스를 생성할 때 지정한 네트워크 수에 의해 결정됩니다. Tenant Port와 Ironic Port의 연관 관계는 Port Group의 수와 각 Port Group의 우선순위에 따라 결정됩니다. 관리자는 베어메탈을 랙에 장착할 때 실제 연결 상황에 따라 동일한 네트워크 평면에 연결된 두 개의 네트워크 카드(Ironic Port)를 하나의 Port Group으로 묶습니다. 두 개의 Ironic Port의 MAC 주소는 다르지만, Port Group의 MAC 주소는 Ironic과 베어메탈 운영 체제에서 일치해야 하므로 하나의 Ironic Port를 주 네트워크 카드로 선택하고 Port Group의 MAC 주소는 주 네트워크 카드의 MAC을 상속합니다. Port Group의 우선순위는 하나의 베어메탈에 여러 Port Group이 있지만 사용자가 하나의 네트워크만 요청하여 베어메탈 인스턴스를 생성할 때 우선순위가 높은 Port Group에만 연결되도록 보장합니다.

1. **nova-compute**는 **Ironic API**를 호출하여 배포 요청을 발행합니다.
2. 각 베어메탈은 프로비저닝 전에 **ironic-api**가 Ring HASH 방식을 사용하여 현재 사용 가능한 **ironic-conductor** 서비스 중 하나를 선택하여 해당 베어메탈의 배포 작업을 담당합니다. 이 **ironic-conductor** 서비스의 프로비저닝 네트워크 설정이 베어메탈이 특정 프로비저닝 네트워크에 연결되도록 결정합니다.
3. **ironic-conductor**는 베어메탈의 주 네트워크 카드의 **LLDP** 정보를 **Provision Port**에 업데이트합니다. 이 단계는 SDN 측면에서 **LLDP**에 대응하는 스위치 포트로 전송 규칙을 배포해야 함을 의미하며, 즉 해당 스위치 포트를 프로비저닝 네트워크에 추가하는 것입니다.
4. 배포가 완료되면, **ironic-conductor**는 임시 **Provision Port**를 삭제합니다.
5. **ironic-conductor**는 동일한 Port Group에 속하는 두 개의 **Ironic Port**의 **LLDP** 정보를 **Tenant**

Port에 업데이트합니다. 이 단계는 SDN 측면에서 두 개의 **LLDP** 정보에 대응하는 두 개의 스위치 포트로 전송 규칙을 배포해야 함을 의미하며, 즉 이 두 스위치 포트를 테넌트 네트워크에 추가하는 것입니다.

참조링크

- https://www.cnblogs.com/jmilkfan-fanguiju/p/11825061.html
- linuxconjapan2015-shiina.pdf

From:
https://atl.kr/dokuwiki/ - **AllThatLinux!**

Permanent link:
https://atl.kr/dokuwiki/doku.php/ironic_%EB%B2%A0%EC%96%B4%EB%A9%94%ED%83%88_%EA%B4%80%EB%A6%AC_%EC%84%9C%EB%B9%84%EC%8A%A4%EC%9D%98_%EB%84%A4%ED%8A%B8%EC%9B%8C%ED%81%AC_%EB%AA%A8%EB%8D%B8

Last update: **2024/10/11 06:55**