本文主要参考 [Overview of functionality and components](https://github.com/openvswitch/ovs/blob/master/README.md) 以及 [Frequently Asked Questions](https://github.com/openvswitch/ovs/blob/master/FAQ.md) 以及结合自己的理解。

http://sdnhub.cn/index.php/openv-switch-full-guide/

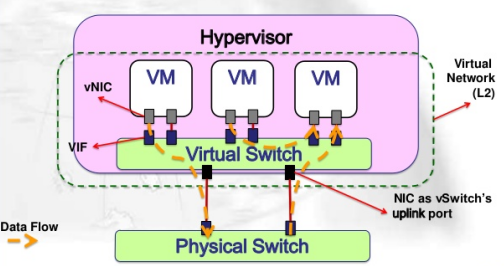
1 什么是OpenvSwitch

OpenvSwitch，简称OVS是一个虚拟交换软件，主要用于虚拟机VM环境，作为一个虚拟交换机，支持Xen/XenServer, KVM, and VirtualBox多种虚拟化技术。OpenvSwitch还支持多个物理机的分布式环境。

在这种某一台物理机器的虚拟化环境中，一个虚拟交换机（vswitch）主要有如下两个作用：

* 传递虚拟机VM之间的流量
* 以及实现VM和外界网络的通信。

如下图所示：

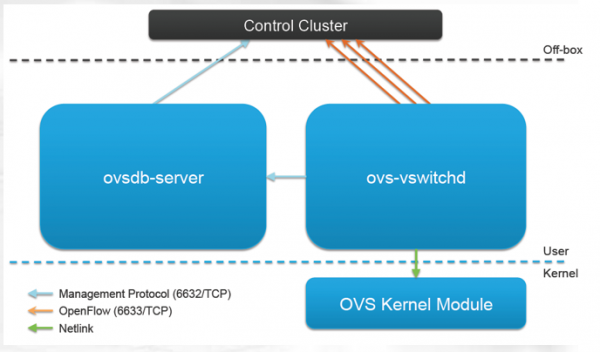
[](http://sdnhub.cn/wp-content/uploads/2015/05/openvswitch-intro.png)

整个OVS代码用C写的。目前有以下功能：

* Standard 802.1Q VLAN model with trunk and access ports
* NIC bonding with or without LACP on upstream switch
* NetFlow, sFlow(R), and mirroring for increased visibility
* QoS (Quality of Service) configuration, plus policing
* GRE, GRE over IPSEC, VXLAN, and LISP tunneling
* 802.1ag connectivity fault management
* OpenFlow 1.0 plus numerous extensions
* Transactional configuration database with C and Python bindings
* High-performance forwarding using a Linux kernel module

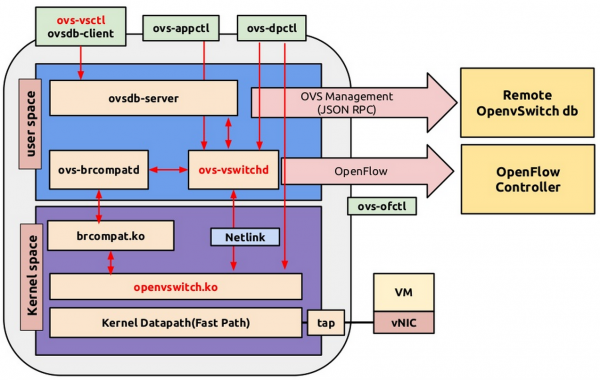
2 OpenvSwitch的组成

OVS的核心组件包括 ovsdb-server，ovs-vswitchd，ovs kernel module。如下图所示：

[](http://sdnhub.cn/wp-content/uploads/2015/05/openvswitch-main-components.png)

运行原理：

内核模块实现了多个“数据路径（DataPath）”（类似于网桥），每个都可以有多个“vports”（类似于桥内的端口）。每个数据路径也通过关联流表（flow table）来设置操作，而这些流表中的流都是用户空间在报文头和元数据的基础上映射的关键信息，一般的操作都是将数据包转发到另一个vport。当一个数据包到达一个vport，内核模块所做的处理是提取其流的关键信息并在流表中查找这些关键信息。当有一个匹配的流时它执行对应的操作。如果没有匹配，它会将数据包送到用户空间的处理队列中（作为处理的一部分，用户空间可能会设置一个流用于以后遇到相同类型的数据包可以在内核中执行操作）。细节如下图所示：

[](http://sdnhub.cn/wp-content/uploads/2015/05/openvswitch-details.png)

除了核心组件，还包括一些管理工具，详细介绍如下：

* ovs-vswitchd：OVS守护进程是OVS的核心部件，实现交换功能，和Linux内核兼容模块一起，实现基于流的交换（flow-based switching）。它和上层 controller 通信遵从 OPENFLOW 协议，它与 ovsdb-server 通信使用 OVSDB 协议，它和内核模块通过netlink通信，它支持多个独立的 datapath（网桥），它通过更改flow table 实现了绑定和VLAN等功能。
* ovsdb-server：OVS轻量级的数据库服务器，用于整个OVS的配置信息，包括接口，交换内容，VLAN 等等。ovs-vswitchd 根据数据库中的配置信息工作。它于 manager 和 ovs-vswitchd 交换信息使用了OVSDB(JSON-RPC)的方式。
* ovs-dpctl：一个工具，用来配置交换机内核模块，可以控制转发规则。
* ovs-vsctl：主要是获取或者更改ovs-vswitchd的配置信息，此工具操作的时候会更新ovsdb-server中的数据库。
* ovs-appctl：主要是向OVS守护进程发送命令的，一般用不上。 a utility that sends commands to running Open vSwitch daemons (ovs-vswitchd)
* ovsdbmonitor：GUI工具来显示ovsdb-server中数据信息。（Ubuntu下是可以使用apt-get安装，可以远程获取OVS数据库和OpenFlow的流表）
* ovs-controller：一个简单的OpenFlow控制器
* ovs-ofctl：用来控制OVS作为OpenFlow交换机工作时候的流表内容。
* ovs-pki：OpenFlow交换机创建和管理公钥框架；
* ovs-tcpundump：tcpdump的补丁，解析OpenFlow的消息；
* brocompat.ko : Linux bridge compatibility module
* openvswitch.ko : Open vSwitch switching datapath

一些用用的OVS命令示例：

sudo ovs-vsctl show

sudo ovs-vsctl add-br mybridge

sudo ovs-vsctl del-br mybridge

sudo ovs-vsctl add-port mybridge port-name

sudo ovs-vsctl del-port mybridge port-name

sudo ovs-vsctl list Bridge/Port/Interface/...

sudo ovs-appctl fdb/show mybridge

sudo ovs-ofctl show mybridge

sudo ovs-ofctl dump-flows mybridge

sudo ovs-ofctl add-flow mybridge dl\_src=02:a2:a2:a2:a2:a2,dl\_dst=02:b2:b2:b2:b2:b2,in\_port=2,dl\_type=0x0800,nw\_src=10.0.0.1,nw\_dst=10.0.0.2,actions=output:6

sudo ovs-ofctl del-flows mybridge dl\_src=02:a2:a2:a2:a2:a2,dl\_dst=02:b2:b2:b2:b2:b2,in\_port=2,dl\_type=0x0800,nw\_src=10.0.0.1,nw\_dst=10.0.0.2

sudo ovs-ofctl add-flow dp0 in\_port=2,actions=output:6

# This will delete all the flow entries in the flow table

sudo ovs-ofctl del-flows mybridge

3 OpenvSwitch和其他vswitch

这里其他的vswitch,包括VMware vNetwork distributed switch以及思科的Cisco Nexus 1000V。

VMware vNetwork distributed switch以及思科的Cisco Nexus 1000V这种虚拟交换机提供的是一个集中式的控制方式，。而OVS则是一个独立的vswitch，他运行在每个实现虚拟化的物理机器上，并提供远程管理。

OVS提供了两种在虚拟化环境中远程管理的协议：

* 一个是OpenFlow,通过流表来管理交换机的行为，
* 一个是OVSDB management protocol，用来暴露sietch的port状态。

4 概念及工作流程

4.1 vswitch、Bridge、Datapath

在网络中，交换机和桥都是同一个概念，OVS实现了一个虚拟机的以太交换机，换句话说，OVS也就是实现了一个以太桥。那么，在OVS中，给一个交换机，或者说一个桥，用了一个专业的名词，叫做DataPath！

要了解OVS如何工作，首先需要知道桥的概念:

网桥也叫做桥接器，连接两个局域网的设备，网桥工作在数据链路层，将两个LAN连接，根据MAC地址来转发帧，可以看成一个“低层的路由器”（路由器工作在网络层，根据IP地质进行转发）。

网桥的工作原理

网桥处理包遵循以下几条规则：

* 在一个接口上接收到的包不会再往那个接口上发送此包。
* 每个接收到的包都要学习其源MAC地址。
* 如果数据包是多播或者广播包（通过2层MAC地址确定）则要向接收端口以外的所有端口转发，如果上层协议感兴趣，则还会递交上层处理。
* 如果数据包的地址不能再CAM表中找到，则向接收端口以外的其他端口转发。
* 如果CAM表中能找到，则转发给相应端口，如果发送和接收都是统一端口，则不发送。

注意，网桥是以混杂模式工作的。关于网桥更多，请查阅相关资料。

OVS中的bridge

上面，说到，一个桥就是一个交换机。例如，在OVS中:

root@localhost:~# ovs-vsctl add-br br0

root@localhost:~# ifconfig br0

br0 Link encap:Ethernet HWaddr 1a:09:56:ea:0b:49

inet6 addr: fe80::1809:56ff:feea:b49/64 Scope:Link

UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

RX packets:1584 errors:0 dropped:0 overruns:0 frame:0

TX packets:6 errors:0 dropped:0 overruns:0 carrier:0

collisions:0 txqueuelen:0

RX bytes:316502 (316.5 KB) TX bytes:468 (468.0 B)

当我们创建了一个交换机（网桥）以后，此时网络功能不受影响，但是会产生一个虚拟网卡，名字就是网桥的名称（br-int），之所以会产生一个虚拟网卡，是为了实现接下来的网桥（交换机）功能。有了这个交换机以后，还需要为这个交换机增加端口(port)，一个端口，就是一个物理网卡，当网卡加入到这个交换机之后，其工作方式就和普通交换机的一个端口的工作方式类似了。

root@localhost:~# ovs-vsctl add-port br0 port

这里要特别注意，网卡加入网桥以后，要按照网桥的工作标准工作，那么加入的一个端口就必须是以混杂模式工作，工作在链路层，处理2层的帧，所以这个port就不需要配置IP了。（你没见过哪个交换的端口有IP的吧）

那么接下来你可能会问，通常的交换机不都是有一个管理接口，可以telnet到交换机上进行配置吧，那么在OVS中创建的虚拟交换机有木有这种呢，有的！上面既然创建交换机brname的时候产生了一个虚拟网口 br-int，那么，你给这个虚拟网卡配置了IP以后，就相当于给交换机的管理接口配置了IP，此时一个正常的虚拟交换机就搞定了。

root@localhost:~# ip address add 192.168.1.1/24 dev br0

最后，我们来看看一个br的具体信息：

root@localhost:~# ovs-vsctl show

bc12c8d2-6900-42dd-9c1c-30e8ecb99a1b

Bridge "br0"

Port "eth0"

Interface "eth0"

Port "br0"

Interface "br0"

type: internal

ovs\_version: "1.4.0+build0"

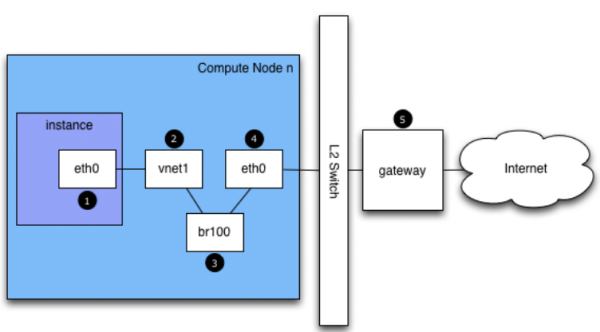
首先，这里显示了一个名为br0的桥（交换机），这个交换机有两个接口,一个是eth0，一个是br0，上面说到，创建桥的时候会创建一个和桥名字一样的接口，并自动作为该桥的一个端口，那么这个虚拟接口的作用，一方面是可以作为交换机的管理端口，另一方面也是基于这个虚拟接口，实现了桥的功能。

#### OpenvSwitch的典型工作流程

这一部分以一个简单的例子，说明在虚拟化环境中OpenvSwitch的典型工作流程。

前面已经说到，OVS主要是用来在虚拟化环境中。实现虚拟机之间通信以及一个虚拟机和外网之间通信，如下是一个典型的结构图：

那么，通常情况下的工作流程如下：

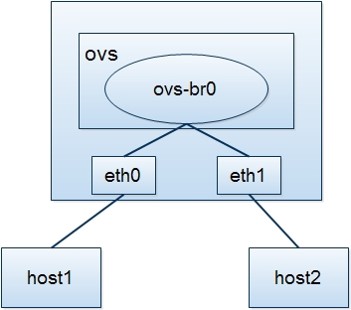
[](http://sdnhub.cn/wp-content/uploads/2015/05/logical_model.png)

1. VM实例 instance 产生一个数据包并发送至实例内的虚拟网络接口 VNIC，图中就是 instance 中的 eth0.
2. 这个数据包会传送到物理机上的VNIC接口，如图就是vnet接口。
3. 数据包从 vnet NIC 出来，到达桥（虚拟交换机） br100 上.
4. 数据包经过交换机的处理，从物理节点上的物理接口发出，如图中物理机上的 eth0 .
5. 数据包从 eth0 出去的时候，是按照物理节点上的路由以及默认网关操作的，这个时候该数据包其实已经不受你的控制了。

一般 L2 switch 连接 eth0 的这个口是一个 trunk 口, 因为虚拟机对应的 VNET 往往会设置 VLAN TAG, 可以通过对虚拟机对应的 vnet 打 VALN TAG 来控制虚拟机的网络广播域. 如果跑多个虚拟机的话, 多个虚拟机对应的 vnet 可以设置不同的 vlan tag, 那么这些虚拟机的数据包从 eth0(4)出去的时候, 会带上TAG标记. 这样也就必须是 trunk 口才行.

6 OVS简单应用实例

6.1 创建物理机到物理机的网络拓扑

[](http://sdnhub.cn/wp-content/uploads/2015/05/example-topo.jpg)

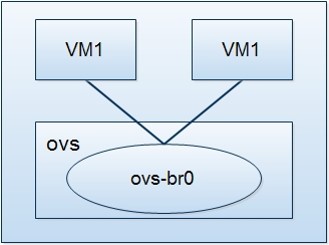
root@localhost:~# ovs-vsctl add-br br0

root@localhost:~# ovs-vsctl add-port br0 eth0

root@localhost:~# ovs-vsctl add-port br0 eth1

6.2 创建虚拟机到虚拟机的网络拓扑

使用KVM虚拟化。

[](http://sdnhub.cn/wp-content/uploads/2015/05/example-topo2.jpg)

root@localhost:~# ovs-vsctl add-br br0

root@localhost:~# cat /etc/ovs-ifup

#!/bin/sh

switch='br0'

/sbin/ifconfig $1 0.0.0.0 up

ovs-vsctl add-port ${switch} $1

root@localhost:~# cat /etc/ovs-ifdown

#!/bin/sh

switch= 'br0'

ovs-vsctl del-port ${sw/sbin/ifconfig $1 0.0.0.0 down

itch} $1

root@localhost:~# kvm -m 512 -net nic,macaddr=00:11:22:33:44:55 -net \

tap,script=/etc/ovs-ifup,downscript=/etc/ovs-ifdown-drive \

file=/path/to/disk-image,boot=on          root@localhost:~# kvm -m 512 -net nic,macaddr=11:22:33:44:55:66 -net \      tap,script=/etc/ovs-ifup,downscript=/etc/ovs-ifdown-drive \

file=/path/to/disk-image,boot=on

7.Misc

7.1 查表

ovs-vsctl list bridge ovs-br

7.2 关于 Brdige 及 Port

创建 Brdige

ovs-vsctl add-br ovs-br

在 ovs-br 上添加 interface

ovs-vsctl add-port ovs-br eth0

(1) + (2) 的可以合并为：

ovs−vsctl add−br ovs-br -- add−port ovs-br eth0

删除 Bridge

ovs-vsctl del-br ovs-br #如果不存在, 会有error log

ovs-vsctl --if-exists del-br ovs-br

更改 ofport (openflow port number)为100：

ovs-vsctl add-port ovs-br eth0 -- set Interface eth0 ofport\_request=100

设置 port 为 internal

ovs-vsctl set Interface eth0 type=internal

7.3 关于Controller

设置 Controller

ovs-vsctl set-controller ovs-br tcp:1.2.3.4:6633

设置 multi controller   
ovs-vsctl set-controller ovs-br tcp:1.2.3.4:6633 tcp:5.6.7.8:6633

查询 Controller

ovs-vsctl show

如果有成功连到 controller 则提示 is\_connected:true, 反之未连上：

ovs-vsctl get-controller ovs-br

移除 Controller   
ovs-vsctl del-controller ovs-br

7.4 关于 STP (Spanning Tree Protocol)

开启 STP

ovs-vsctl set bridge ovs-br stp\_enable=true

关闭 STP

ovs-vsctl set bridge ovs-br stp\_enable=false

查询 STP 配置信息

ovs-vsctl get bridge ovs-br stp\_enable

设置 Priority

ovs−vsctl set bridge br0 other\_config:stp-priority=0x7800

设置 Cost   
ovs−vsctl set port eth0 other\_config:stp-path-cost=10

移除 STP 设置

ovs−vsctl clear bridge ovs-br other\_config

7.5 关于 Openflow Version

支持 OpenFlow Version 1.3

ovs-vsctl set bridge ovs-br protocols=OpenFlow13

支持 OpenFlow Version 1.3 1.2

ovs-vsctl set bridge ovs-br protocols=OpenFlow12,OpenFlow13

移除 OpenFlow 支持设置

ovs-vsctl clear bridge ovs-br protocols

7.6 关于 VLAN

设置 VLAN tag

ovs-vsctl add-port ovs-br vlan3 tag=3 -- set interface vlan3 type=internal

移除 VLAN

ovs-vsctl del-port ovs-br vlan3

查询 VLAN

ovs-vsctl show

ifconfig vlan3

设置 Vlan trunk   
ovs-vsctl add-port ovs-br eth0 trunk=3,4,5,6

设置已 add 的 port 为 access port, vlan id 9

ovs-vsctl set port eth0 tag=9

ovs-ofctl add-flow 设置 vlan 100

ovs-ofctl add-flow ovs-br in\_port=1,dl\_vlan=0xffff,actions=mod\_vlan\_vid:100,output:3

ovs-ofctl add-flow ovs-br in\_port=1,dl\_vlan=0xffff,actions=push\_vlan:0x8100,set\_field:100-\>vlan\_vid,output:3

ovs-ofctl add-flow 拿掉 vlan tag

ovs-ofctl add-flow ovs1 in\_port=3,dl\_vlan=100,actions=strip\_vlan,output:1

two\_vlan example

ovs-ofctl add-flow pop-vlan

ovs-ofctl add-flow ovs-br in\_port=3,dl\_vlan=0xffff,actions=pop\_vlan,output:1

7.7 关于 GRE Tunnel

设置 GRE tunnel

ovs−vsctl add−port ovs-br ovs-gre -- set interface ovs-gre type=gre options:remote\_ip=1.2.3.4

查询 GRE Tunnel

ovs-vsctl show

7.8 关于 Dump flows

Dumps OpenFlow flows 不含 hidden flows (常用)

ovs-ofctl dump-flows ovs-br

Dumps OpenFlow flows 包含 hidden flows

ovs-appctl bridge/dump-flows ovs-br

Dump 特定 bridge 的 datapath flows 不論任何 type

ovs-appctl dpif/dump-flows ovs-br

Dump 在 Linux kernel 裡的 datapath flow table (常用)

ovs-dpctl dump-flows [dp]

Top like behavior for ovs-dpctl dump-flows

ovs-dpctl-top

7.9 XenServer 开启 OpenvSwitch 方式

检查是否启动openvswitch服务：

service openvswitch status

启动服务

xe-switch-network-backend openvswitch

关闭服务

xe-switch-network-backend bridge

7.10 关于 Log

查询 log level list

ovs-appctl vlog/list

设置 log level (以 stp 设置 file 为 dbg level 为例)

ovs-appctl vlog/set stp:file:dbg

ovs-appctl vlog/set {module name}:{console, syslog, file}:{off, emer, err, warn, info, dbg}

7.11 关于 Fallback

Controller connection: false 的时候, 会自动调成 legacy switch mode

ovs-vsctl set-fail-mode ovs-br standalone

无论 Controller connection status 为何, 都必须通过 OpenFlow 进行网络行为 (default)

ovs-vsctl set-fail-mode ovs-br secure

移除

ovs-vsctl del-fail-mode ovs-br

查询

ovs-vsctl get-fail-mode ovs-br

7.12 关于 sFlow

查询

ovs-vsctl list sflow

新增

Set sFlow 缺

刪除

ovs-vsctl -- clear Bridge ovs-br sflow

7.13关于 NetFlow

查询

ovs-vsctl list netflow

新增

Set NetFlow 缺

刪除

ovs-vsctl -- clear Bridge ovs-br netflow

7.14 设置 Out-of-band 和 in-band

查询

ovs-vsctl get controller ovs-br connection-mode

Out-of-band

ovs-vsctl set controller ovs-br connection-mode=out-of-band

In-band (default)

ovs-vsctl set controller ovs-br connection-mode=in-band

移除 hidden flow

ovs-vsctl set bridge br0 other-config:disable-in-band=true

7.15 关于 ssl

查询

ovs-vsctl get-ssl

设置

ovs-vsctl set-ssl sc-privkey.pem sc-cert.pem cacert.pem

OpenvSwitch Lab 6$ TLS SSL : http://roan.logdown.com/posts/208707-openvswitch-lab-6-ssl

刪除

ovs-vsctl del-ssl

7.16 关于 SPAN

详细设置

ovs-vsctl add-br ovs-br

ovs-vsctl add-port ovs-br eth0

ovs-vsctl add-port ovs-br eth1

ovs-vsctl add-port ovs-br tap0 \

-- --id=@p get port tap0 \

-- --id=@m create mirror name=m0 select-all=true output-port=@p \

-- set bridge ovs-br mirrors=@m

将 ovs-br 上 add-port {eth0,eth1} mirror 至 tap0

刪除

ovs-vsctl clear bridge ovs-br mirrors # 關於 Table

查 table ovs-ofctl dump-tables ovs-br

7.17 关于 Group Table

参考 [hwchiu – Multipath routing with Group table at mininet](http://hwchiu.logdown.com/posts/207387-multipath-routing-with-group-table-at-mininet)

建立 Group id 及对应的 bucket

ovs-ofctl -O OpenFlow13 add-group ovs-br group\_id=5566,type=select,bucket=output:1,bucket=output:2,bucket=output:3

type 共有 All, Select, Indirect, FastFailover, 详细规格：<http://flowgrammable.org/sdn/openflow/message-layer/groupmod/#GroupMod_1.3>

使用 Group Table

ovs-ofctl -O OpenFlow13 add-flow ovs-br in\_port=4,actions=group:5566

7.18 关于 VXLAN

参考 [rascov – Bridge Remote Mininets using VXLAN](http://rascov.logdown.com/posts/230635-bridge-remote-networks-using-vxlan)

建立 VXLAN Network ID (VNI) 和指定的 OpenFlow port number, eg: VNI=5566, OF\_PORT=9

ovs-vsctl set interface vxlan type=vxlan option:remote\_ip=x.x.x.x option:key=5566 ofport\_request=9

VNI flow by flow

ovs-vsctl set interface vxlan type=vxlan option:remote\_ip=140.113.215.200 option:key=flow ofport\_request=9

设置 VXLAN tunnel id

ovs-ofctl add-flow ovs-br in\_port=1,actions=set\_field:5566->tun\_id,output:2

ovs-ofctl add-flow s1 in\_port=2,tun\_id=5566,actions=output:1

7.19 关于 OVSDB Manager

参考 [OVSDB Integration:Mininet OVSDB Tutorial](https://wiki.opendaylight.org/view/OVSDB_Integration:Mininet_OVSDB_Tutorial)

Active Listener 设置

ovs-vsctl set-manager tcp:1.2.3.4:6640

Passive Listener 设置

ovs-vsctl set-manager ptcp:6640

7.20 OpenFlow Trace

Generate pakcet trace

ovs-appctl ofproto/trace ovs-br in\_port=1,dl\_src=00:00:00:00:00:01,dl\_dst=00:00:00:00:00:02 -generate

7.21 其它

查询 OpenvSwitch 版本

ovs-ofctl -V

查询指令历史记录

ovsdb-tool show-log [-mmm]

[参考]

1. <http://git.openvswitch.org/cgi-bin/gitweb.cgi?p=openvswitch;a=blob_plain;f=FAQ;hb=HEAD>
2. <http://git.openvswitch.org/cgi-bin/gitweb.cgi?p=openvswitch;a=blob_plain;f=README;hb=HEAD>
3. <http://openvswitch.org/support/config-cookbooks/vlan-configuration-cookbook/>
4. <http://roan.logdown.com/posts/191801-set-openvswitch>
5. <http://roan.logdown.com/posts/208707-openvswitch-lab-6-ssl>
6. <http://flowgrammable.org/sdn/openflow/message-layer/groupmod/#GroupMod_1.3>
7. <http://www.areteix.net/blog/2013/08/network-flow-monitoring-with-open-vswitch/>
8. <http://blog.scottlowe.org/2014/11/21/removing-ovs-configuration-settings/>