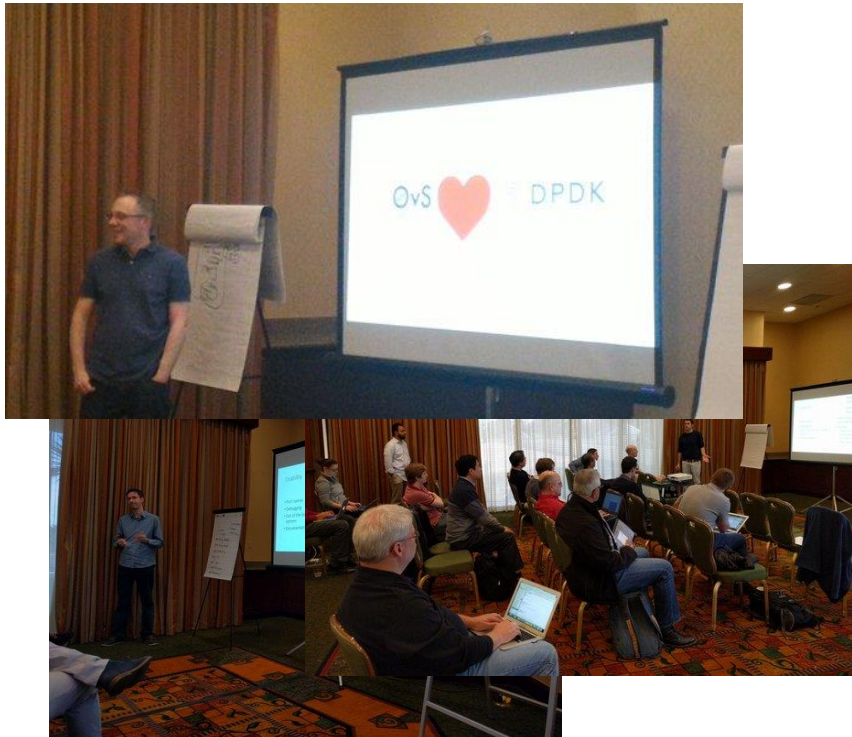


Aaron Conole, Red Hat
Robin Giller, Intel
Bhanuprakash Bodireddy, Intel

OvS-DPDK Usability Improvements for Real-World Applications

This time last year...

In a conference room not too far away



Branch: master ovs / INSTALL.DPDK.rst Find file

mvpolito doc: fix bad link to dpdk advance installation guide f799f7a 2

2 contributors

605 lines (445 sloc) 19.9 KB Raw Blame History

Open vSwitch with DPDK

This document describes how to build and install Open vSwitch using a DPDK datapath. Open vSwitch can use the DPDK library to operate entirely in userspace.

Warning

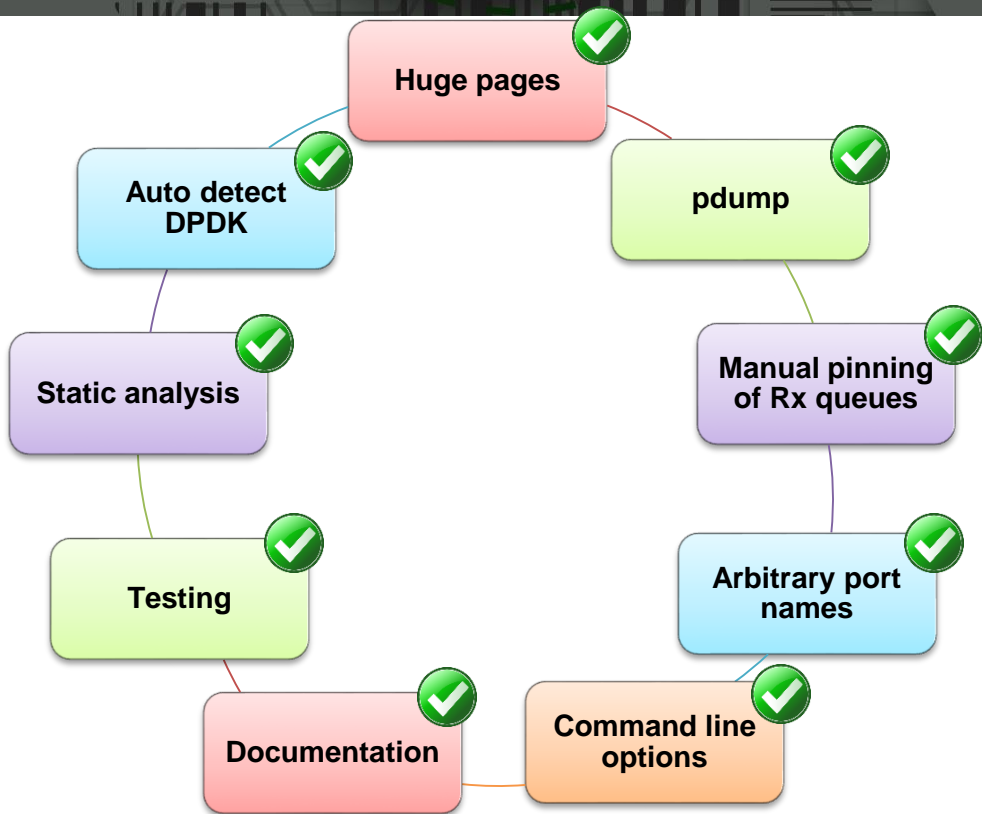
The DPDK support of Open vSwitch is considered 'experimental'.

Build requirements

In addition to the requirements described in the [installation guide](#), building Open vSwitch with DPDK will require the following:

- DPDK 16.07
- A [DPDK supported NIC](#)
 - Only required when physical ports are in use
- A suitable kernel

Usability checklist



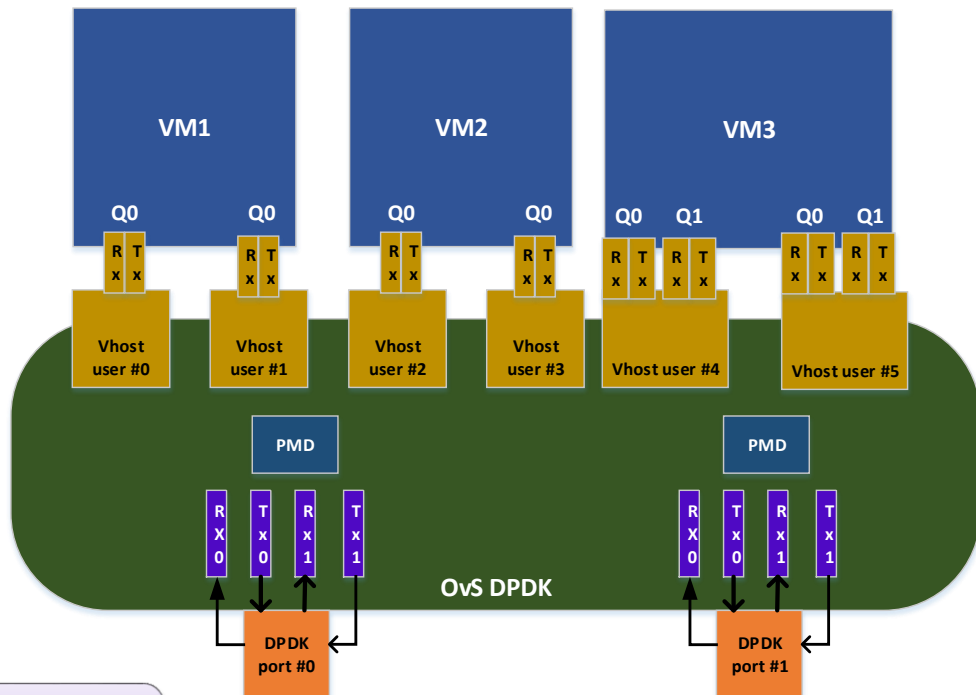
The DPDK support of OvS is considered
'excellent'

Huge pages

- Huge pages need kernel and HW support
 - CONFIG_HUGETLBFS should be enabled
 - Architectures support multiple pages(4k, 8k, 64k, 256K, 1M, 4M, 16M, 256M, 1G)
 - Cpu flags(**pse** – 2M, **pdpe1gb** – 1G huge pages)
- Gigantic pages (1GB pages)
- Persistent vs run-time allocation of Huge pages
- Performance

Huge pages

Manual pinning of rx queues



Manual pinning
of Rx queues

```
pmd thread numa_id 0 core_id 4:  
isolated : false  
port: dpdk0 queue-id: 0  
port: dpdk1 queue-id: 0  
port: dpdkvhostuser0 queue-id: 0  
port: dpdkvhostuser2 queue-id: 0  
port: dpdkvhostuser4 queue-id: 0  
port: dpdkvhostuser5 queue-id: 1  
pmd thread numa_id 0 core_id 5:  
isolated : false  
port: dpdk0 queue-id: 1  
port: dpdk1 queue-id: 1  
port: dpdkvhostuser1 queue-id: 0  
port: dpdkvhostuser3 queue-id: 0  
port: dpdkvhostuser4 queue-id: 1  
port: dpdkvhostuser5 queue-id: 0
```



```
pmd thread numa_id 0 core_id 4:  
isolated : true  
port: dpdk0 queue-id: 0 1  
port: dpdkvhostuser0 queue-id: 0  
port: dpdkvhostuser3 queue-id: 0  
port: dpdkvhostuser2 queue-id: 0  
port: dpdkvhostuser1 queue-id: 0  
pmd thread numa_id 0 core_id 5:  
isolated : true  
port: dpdkvhostuser4 queue-id: 0 1  
port: dpdkvhostuser5 queue-id: 0 1  
port: dpdk1 queue-id: 0 1
```

Arbitrary port names & Auto detect DPDK library

```
$ ovs-vsctl add-port $BRIDGE dpdk0 – set interface dpdk0 type=dpdk
```

Arbitrary port names

↓

```
$ ovs-vsctl add-port $BRIDGE DPDKRX – set interface DPDKRX type=dpdk options:dpdk-devargs:0000:03:00.1
```

'dpdk-devargs' to indicate which DPDK port to associate with a given OVS port

```
$ export DPDK_BUILD=$DPDK_DIR/$DPDK_TARGET  
$ ./configure --with-dpdk=$DPDK_BUILD
```

Auto detect DPDK

↓

```
$ ./configure --with-dpdk
```

Auto discover DPDK library/headers if present in compiler search paths.

Documentation and static analysis

INSTALL.DPDK.md

- Install DPDK, OvS
- setup OvS

INSTALL.DPDK-ADVANCED.md

- System configuration
- Performance Tuning
- Test cases
- Vhost walkthrough

Documentation

```
Command Line: make
Clang Version: clang version 3.5.0 (tags/RELEASE_350/final)
Date: Wed Nov 2 21:00:59 2016
```

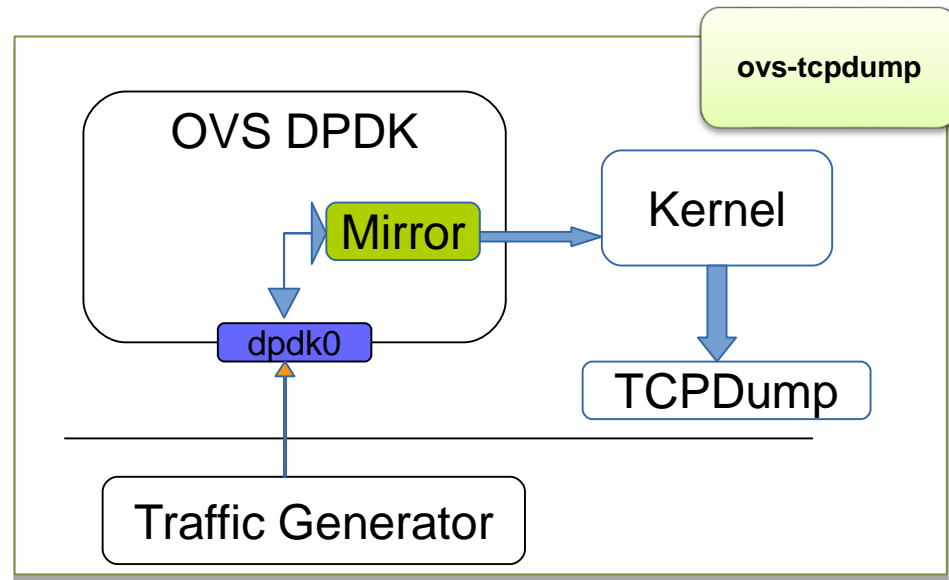
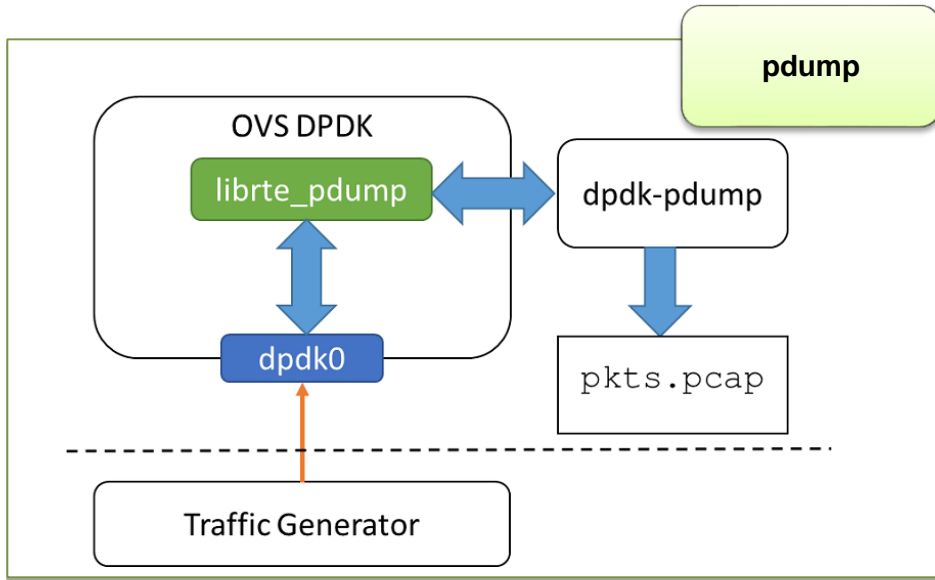
Bug Summary

Bug Type	Quantity	Display?
All Bugs	42	<input checked="" type="checkbox"/>
API		
Argument with 'nonnull' attribute passed null	13	<input checked="" type="checkbox"/>
Dead store		
Dead assignment	1	<input checked="" type="checkbox"/>
Logic error		
Assigned value is garbage or undefined	2	<input checked="" type="checkbox"/>
Branch condition evaluates to a garbage value	1	<input checked="" type="checkbox"/>
Dangerous variable-length array (VLA) declaration	1	<input checked="" type="checkbox"/>

Static analysis

- `./boot.sh`
- `./configure CC=clang/`
- `./configure CC=gcc`
- `make clang-analyze`
- `Scan-view <results dir>`

Packet Tracing – Two approaches



OvS DPDK Start/Stop and library options

```
$ ovs-ctl --no-ovs-vswitchd start  
$ ovs-vsctl --no-wait set Open_vSwitch . other_config:dppk-init=true  
$ ovs-ctl --no-ovsdb-server start
```

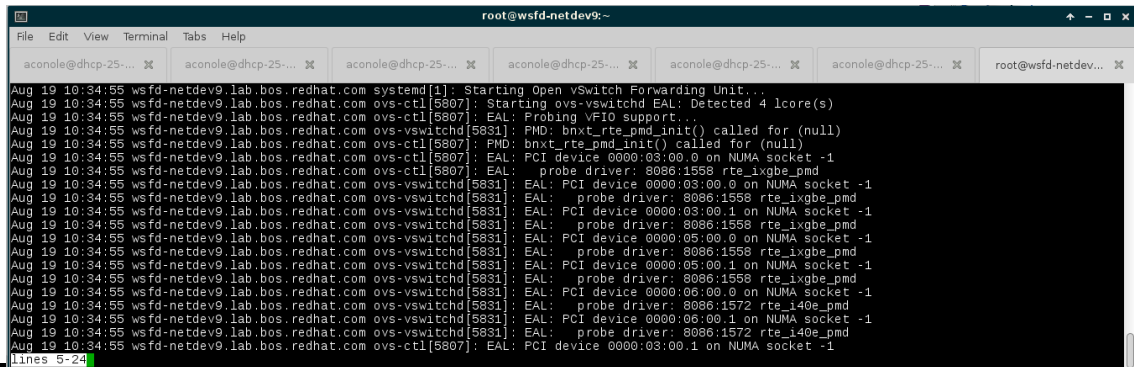
Command line options

(OR)

The RHEL systemd way

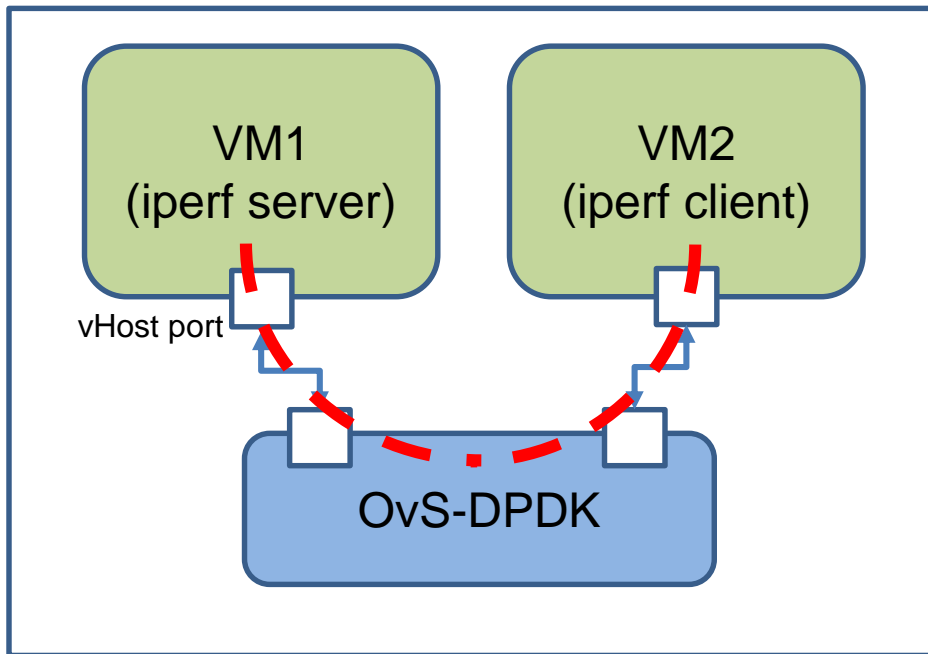
```
$ systemctl start ovssdb-server  
$ ovs-vsctl --no-wait set Open_vSwitch . other_config:dppk-init=true  
$ systemctl start openvswitch
```

Values are set with **'other_config:dppk-extras'** or via a specific database key.



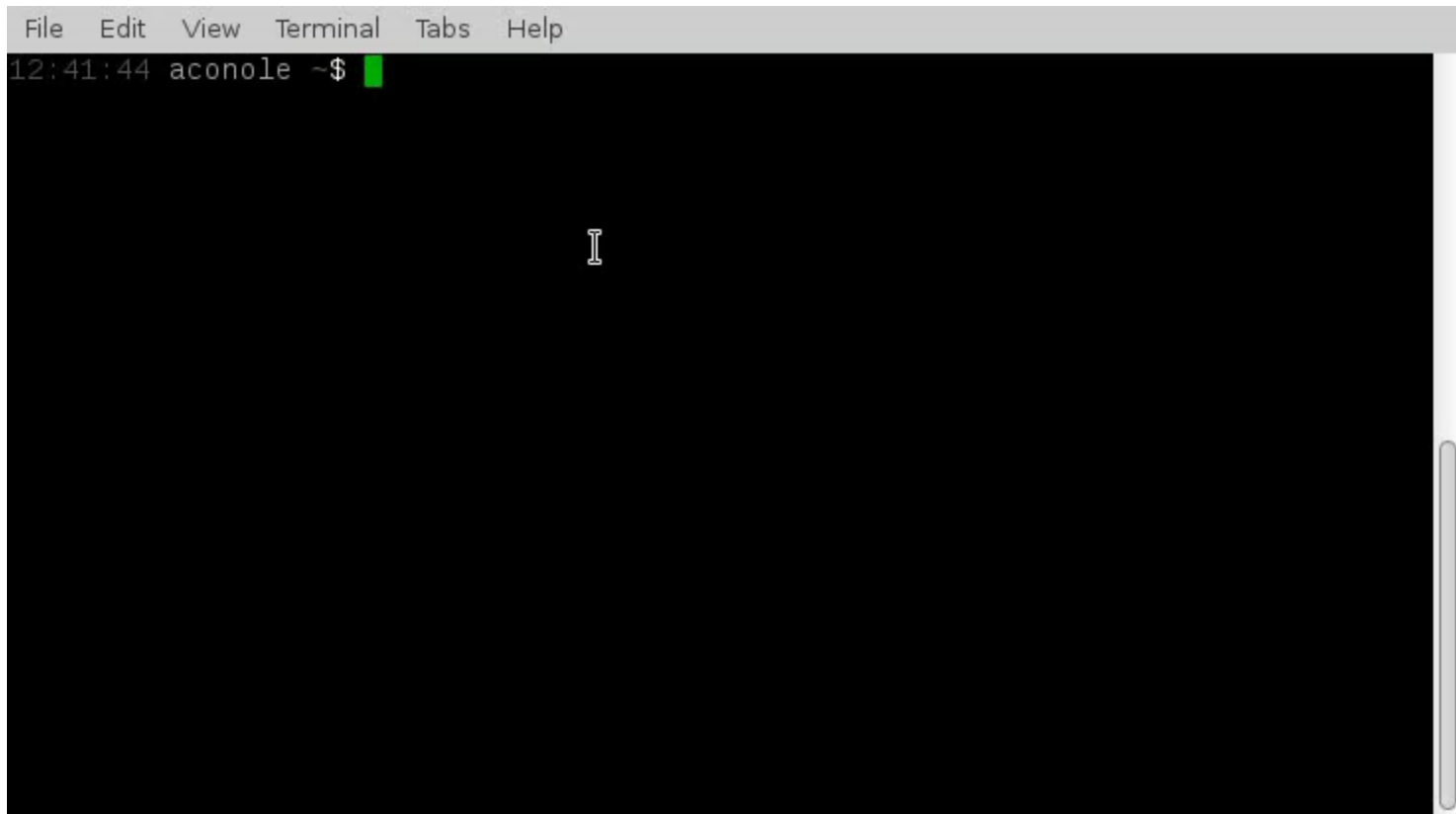
```
root@wsfd-netdev9:~  
File Edit View Terminal Tabs Help  
aconole@dhcp-25... x aconole@dhcp-25... x aconole@dhcp-25... x aconole@dhcp-25... x aconole@dhcp-25... x aconole@dhcp-25... x root@wsfd-netdev... x  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com systemd[1]: Starting Open vSwitch Forwarding Unit...  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: Starting ovs-vswitchd EAL: Detected 4 lcore(s)  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: EAL: Probing VFIO support...  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: PMD: bnxt_rte_pmd_init() called for (null)  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: PMD: bnxt_rte_pmd_init() called for (null)  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: EAL: PCI device 0000:03:00.0 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: EAL: probe driver: 8086:1558 rte_ixgbe_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: PCI device 0000:03:00.0 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: probe driver: 8086:1558 rte_ixgbe_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: PCI device 0000:03:00.1 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: probe driver: 8086:1558 rte_ixgbe_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: PCI device 0000:03:00.1 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: probe driver: 8086:1572 rte_i40e_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: PCI device 0000:06:00.0 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: probe driver: 8086:1572 rte_i40e_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: PCI device 0000:06:00.1 on NUMA socket -1  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-vswitchd[5831]: EAL: probe driver: 8086:1572 rte_i40e_pmd  
Aug 19 10:34:55 wsfd-netdev9.lab.bos.redhat.com ovs-ctl[5807]: EAL: PCI device 0000:03:00.1 on NUMA socket -1  
lines 5-24
```

Demo



- Configuration
 - Host
 - VM
- Send traffic
- Packet dump

Demo



A terminal window with a menu bar containing 'File', 'Edit', 'View', 'Terminal', 'Tabs', and 'Help'. The terminal content shows the time '12:41:44', the user 'aconole', and the shell prompt '~\$' followed by a green cursor. A white cursor is positioned in the center of the terminal area.

```
File Edit View Terminal Tabs Help
12:41:44 aconole ~$ █
I
```

Summary

- Big focus on usability of OvS-DPDK and DPDK
- Many items addressed and merged in upstream project
- OvS-DPDK consumed by multiple Linux distributors and supported by installers
- Further discussion at OvS-DPDK design summit on Wednesday