



## Switchdev – No More SDK

Elad Raz, Jiri Pirko

Netdev 1.1 | 2016

# Switchdev is here to stay

- What?



- Why?

- Where?

100G DEMO!





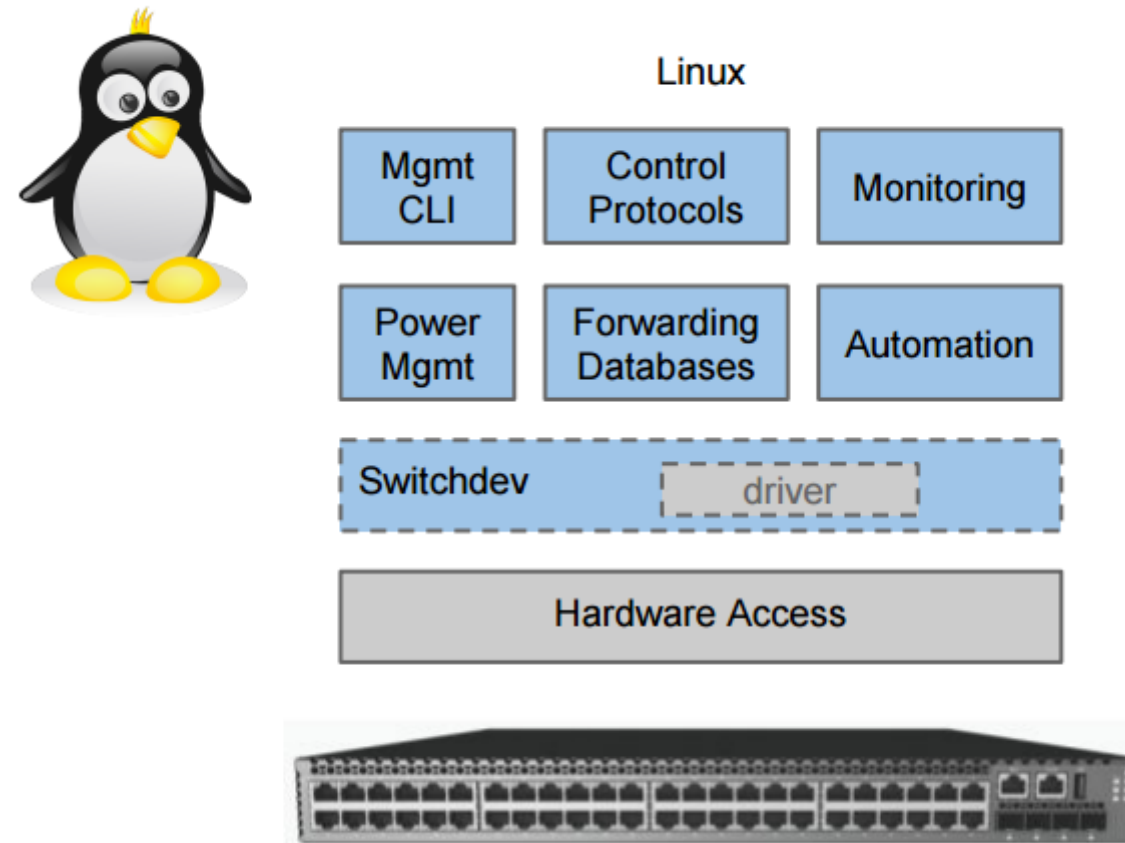
# Switchdev is here to stay

## ■ What?



## ■ Why?

## ■ Where?



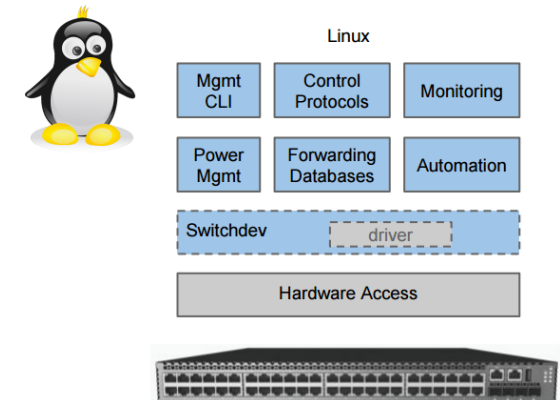
# Switchdev is here to stay

## ■ What?



## 64 net-devices

Doing skb rx/tx, ethtool, port state and ndo offload



# Switchdev is here to stay

## ■ What?

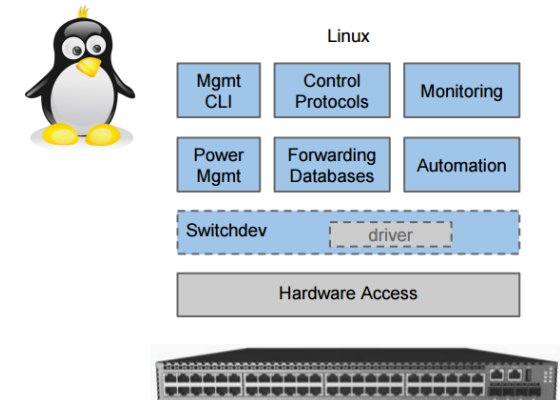


## switchdev\_ops

.1D/.1Q Bridging – VLANs, PVID, FDB, IGMP...

Routing – FIB offload

and more



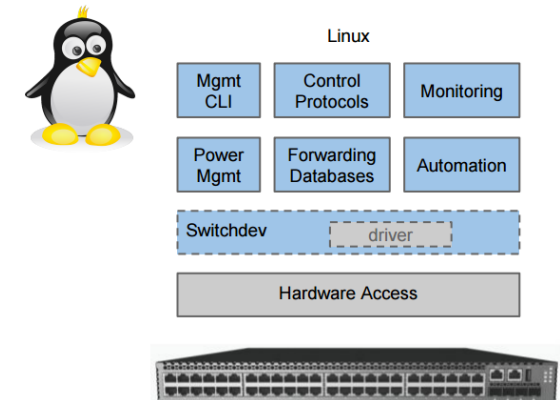
# Switchdev is here to stay

## ■ What?



**In-kernel implementation**  
**NO MORE SDK!**

`drivers/net/ethernet/mellanox/mlxsw`



# Switchdev is here to stay

- What?



- Why?



- Where?



# Switchdev is here to stay

## ■ Why?



Community based  
Open-source  
Standard behavior  
Standard API





# Switchdev is here to stay

- What?



- Why?



- Where?



# Switchdev is here to stay



## Devlink

set port type, port splitter, shared buffers, etc.



■ **Where?**



# Switchdev is here to stay



QoS  
DCBNL interface implementation



■ **Where?**



# Switchdev is here to stay



## L3 offload

fix the current switchdev model, include VRF support



- **Where?**





# Switchdev is here to stay



## Flows

`tc cls_flow + selected acts offload`

■ **Where?**



# Switchdev is here to stay



## (e) BPF?

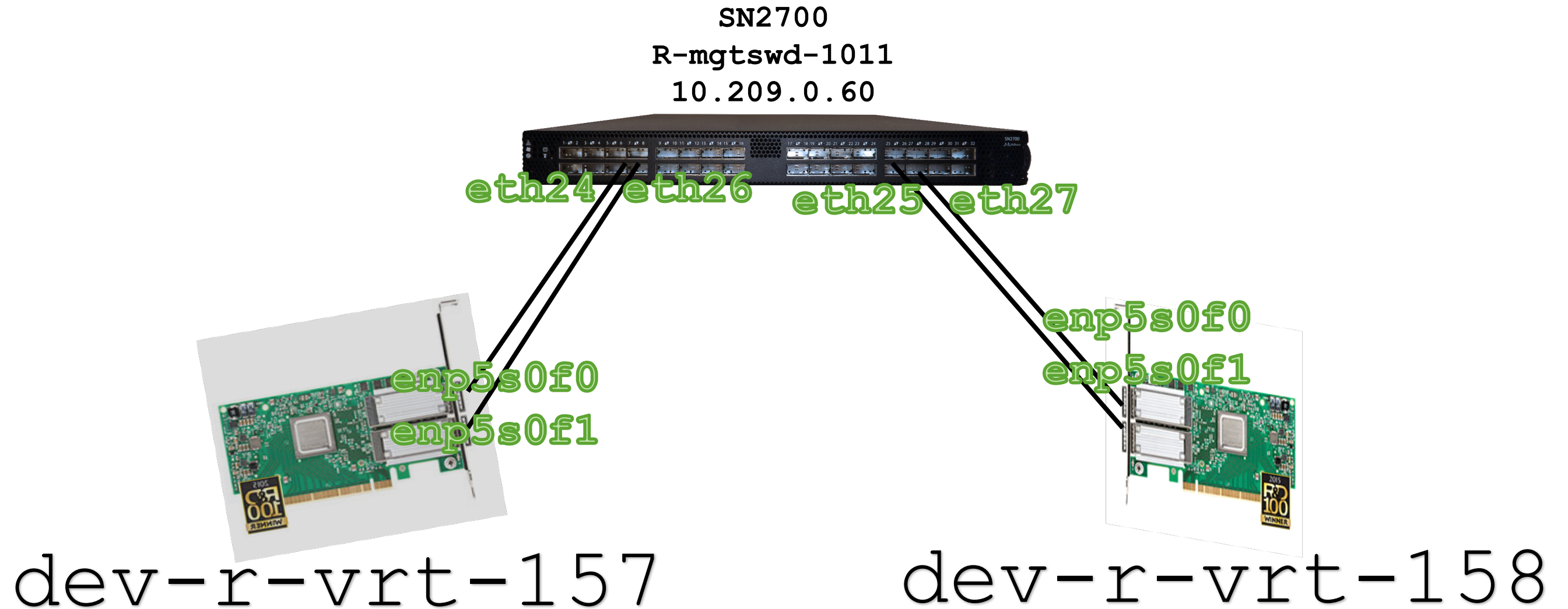
- **Where?**



# Tutorial outlook

- No more talking! 100G demo!
  - Netdevice
  - Bridging
  - 802.1Q bridge
  - 802.1D bridge
  - LAG

# Topology





# Screens

SN2700  
100g switch

server 157  
+ConnectX-4

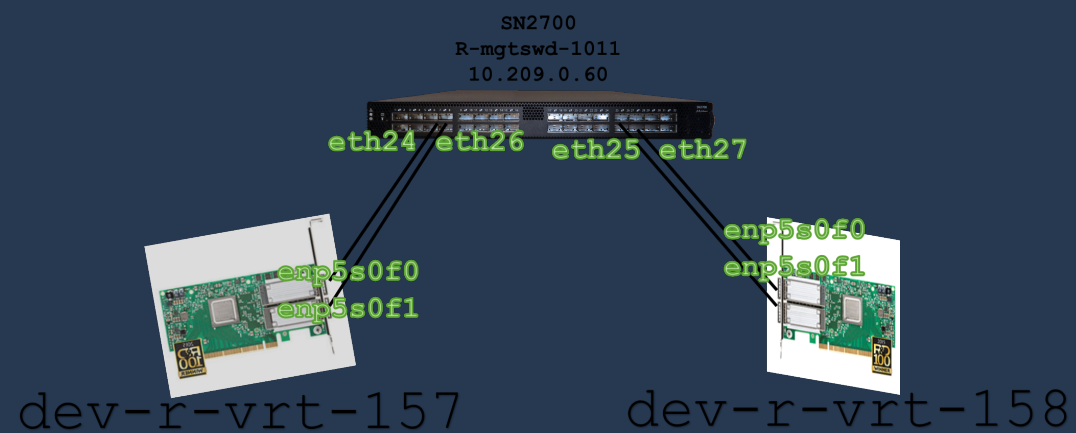
```
Every 1.0s: ip -s link show dev eth24      Sun Feb  7 12:04:11 2016
29: eth24: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel switchid e41d2d4613c0 state UP mode DEFAULT group
default qlen 1000
    link/ether e4:1d:2d:46:13:c0 brd ff:ff:ff:ff:ff:ff
    RX: bytes  packets  errors  dropped  overruns  mcast
         0         0         0         0         0         0
    TX: bytes  packets  errors  dropped  carrier  collisions
       776         8         0         0         0         0
```

SW counters

server 158  
+ConnectX-4

```
Every 1.0s: ethtool -S eth24      Sun Feb  7 12:04:11 2016
NIC statistics:
  a_frames_transmitted_ok: 8
  a_frames_received_ok: 0
  a_frame_check_sequence_error: 0
  a_alignment_errors: 0
  a_octets_transmitted_ok: 680
  a_octets_received_ok: 0
  a_multicast_frames_xmitted_ok: 8
  a_broadcast_frames_xmitted_ok: 0
  a_multicast_frames_received_ok: 0
  a_broadcast_frames_received_ok: 0
  a_in_range_length_errors: 0
```

HW counters



# netdev

SN2700  
100g switch

server 157  
+ConnectX-4

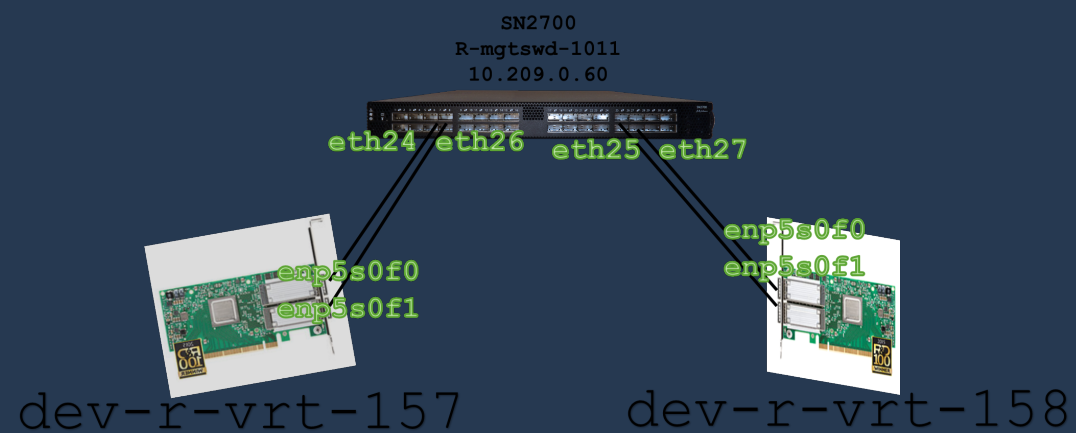
```
Every 1.0s: ip -s link show dev eth24          Sun Feb  7 12:04:11 2016
ip: eth24: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qlen 1000 state UP mode DEFAULT group default qlen 1000
link/ether 01:00:00:00:00:00 brd ff:ff:ff:ff:ff:ff
RX: bytes  packets  errors  dropped  overruns  carrier  missed
0          0          0          0          0          0          0
TX: bytes  packets  errors  dropped  overruns  carrier  missed
74         0          0          0          0          0          0
```

SW counters

```
Every 1.0s: ethtool -S eth24                    Sun Feb  7 12:04:11 2016
NIC statistics:
  rx_bytes_received: 0
  rx_packets_received: 0
  rx_errors_received: 0
  tx_bytes_transmitted: 74
  tx_packets_transmitted: 0
  tx_errors_transmitted: 0
  rx_bytes_received: 0
  rx_packets_received: 0
  rx_errors_received: 0
  tx_bytes_transmitted: 74
  tx_packets_transmitted: 0
  tx_errors_transmitted: 0
  rx_bytes_received: 0
  rx_packets_received: 0
  rx_errors_received: 0
  tx_bytes_transmitted: 74
  tx_packets_transmitted: 0
  tx_errors_transmitted: 0
```

HW counters

server 158  
+ConnectX-4



.1Q bridge

SN2700  
100g switch

server 157  
+ConnectX-4

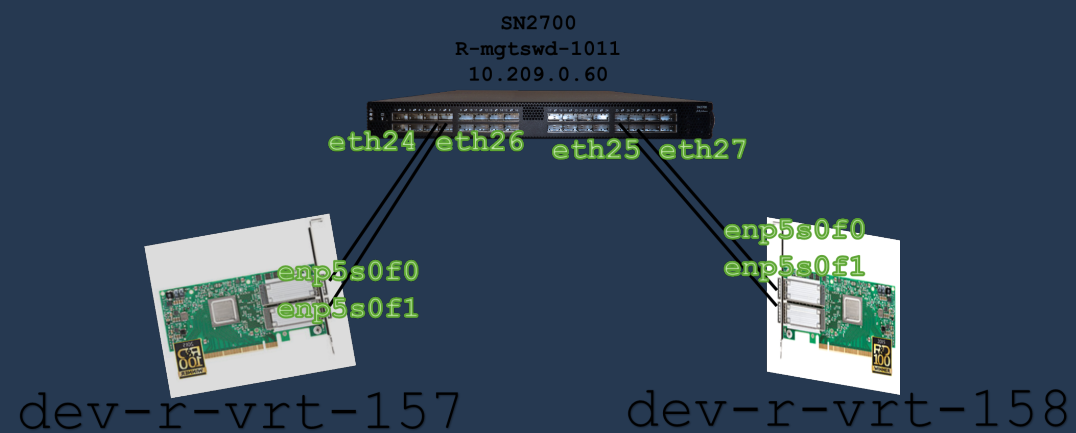
```
Every 1.0s: ip -s link show dev eth24          Sun Feb 7 12:04:11 2016
(2) eth24: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qlen 1000 state UP mode DEFAULT group
default qlen 1000
link/layer eth24:8a:4d:1b:00:00 ff:ff:ff:ff:ff:ff
RX: bytes  packets  errors  dropped  overruns  carrier  missed
0          0          0          0          0          0          0
TX: bytes  packets  errors  dropped  overruns  carrier  missed
74         0          0          0          0          0          0
```

SW counters

```
Every 1.0s: ethtool -S eth24                    Sun Feb 7 12:04:11 2016
NIC statistics:
  rx_bytes: 74
  rx_packets: 0
  rx_errors: 0
  rx_dropped: 0
  tx_bytes: 0
  tx_packets: 0
  tx_errors: 0
  tx_dropped: 0
  multiicast_frames_received: 0
  multiicast_frames_transmitted: 0
  multiicast_frames_received_err: 0
  multiicast_frames_transmitted_err: 0
  rx_fifo_length: 0
```

HW counters

server 158  
+ConnectX-4



# .1D bridge

SN2700  
100g switch

server 157  
+ConnectX-4

```
Every 1.0s: ip -s link show dev eth24 Sun Feb 7 12:04:11 2016
(2) eth24: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qlen 1000 state UP mode DEFAULT group
default qlen 1000
link/layer eth10:2d:4e:11:00:00 ff:ff:ff:ff:ff:ff
RX: bytes packets errors dropped carrier
0 0 0 0 0
TX: bytes packets errors dropped carrier
74 0 0 0 0
```

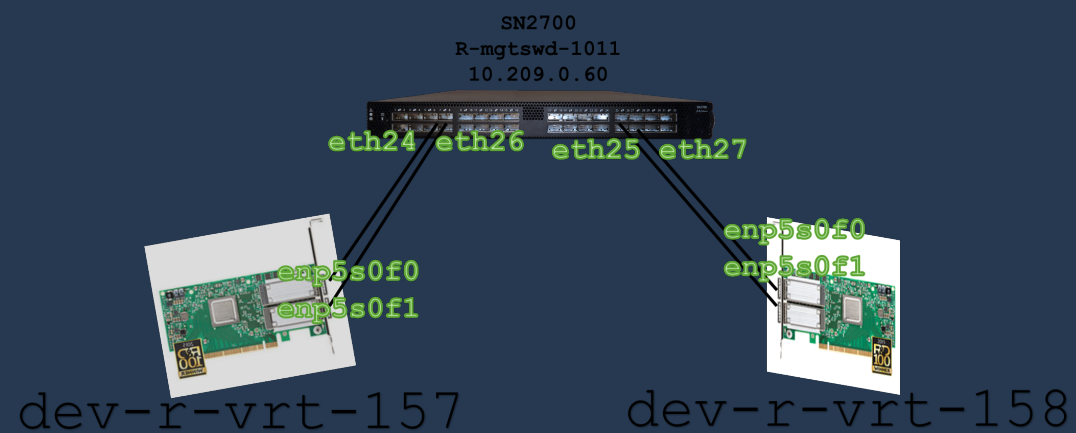
SW counters

server 158  
+ConnectX-4

```
Every 1.0s: cat /sys/class/net/eth24 Sun Feb 7 12:04:11 2016
NIC statistics:
  frames_transmitted_tx: 0
  frames_received_rx: 0
  frame_check_sequence: 0
  alignment_errors: 0
  octets_transmitted_tx: 800
  octets_received_rx: 0
  multibyte_frames_mismatch_rx: 0
  multibyte_frames_mismatch_tx: 0
  multibyte_frames_receive_rx: 0
  multibyte_frames_receive_tx: 0
  rx_ramp_length_error: 0
```

HW counters





# LAG

team and bond

SN2700  
100g switch

server 157  
+ConnectX-4

```
Every 1.0s: ip -s link show dev eth24          Sun Feb  7 12:04:11 2016
(2) eth24: <BRIDGECAST,MULTICAST,UP,LOWER_UP> mtu 1500 qlen 1000 state UP mode DEFAULT group
default qlen 1000
link/layer eth10:2d:4e:11:00:00 ff:ff:ff:ff:ff:ff
RX: bytes  packets  errors  dropped  overruns  carrier  missed
0          0          0          0          0          0          0
TX: bytes  packets  errors  dropped  overruns  carrier  missed
74         0          0          0          0          0          0
```

SW counters

```
Every 1.0s: ethtool -S eth24                    Sun Feb  7 12:04:11 2016
NIC statistics:
  rx_bytes_received: 0
  rx_packets_received: 0
  rx_errors_received: 0
  rx_dropped: 0
  tx_bytes_transmitted: 74
  tx_packets_transmitted: 0
  tx_errors_transmitted: 0
  tx_dropped: 0
  rx_bytes_received: 0
  rx_packets_received: 0
  rx_errors_received: 0
  rx_dropped: 0
  tx_bytes_transmitted: 74
  tx_packets_transmitted: 0
  tx_errors_transmitted: 0
  tx_dropped: 0
```

HW counters

server 158  
+ConnectX-4



# Thank You